

Cooperation can help to get the message across

Those who study natural phenomena that pose risks to society need to work more closely with social scientists to ensure their science is put into practice. Such cooperation could also boost their funding prospects.

If there was not already good reason to doubt the ability of science alone to solve the world's environmental problems, recent events on the Caribbean island of Montserrat would certainly provide it. When nine or more Montserrat residents were engulfed last week by deadly pyroclastic flows from the Soufrière Hills volcano, it was not because of any lack of appreciation, on the part of volcanologists, of how the volcano would behave when it erupted. Instead, the problem lay in the behaviour of those who are said to have ignored evacuation orders in order to tend to their crops and livestock.

There is room, of course, for improvement in the science. In a world of perfectly understood volcanoes, there would be no false alarms. Climactic eruptions would proceed as soon as the population had been evacuated, without displaced residents being left to cool their heels for weeks, suspecting that the scientists had got it wrong. But to yearn for such a volcanological Camelot is to miss the point: the best science in the world is useless if it is not communicated effectively to its potential beneficiaries.

In natural hazards research, effective communication means much more than just translating scientific results — for example both the potential and likely limits of the ability to predict earthquakes (see page 4) — into lay language. Individuals, institutions and communities respond to risk in complex ways that can be studied by behavioural and social scientists. Seismologists are said to have walked into US city halls with maps showing the vulnerability of local structures to likely earthquakes, only to be shown the door — the elected officials simply did not want to know. Such problems can be tackled only within the appropriate legal or political framework.

Social scientists are also needed to assess the full costs of natural disasters; without such an assessment, the cost-effectiveness of potential hazard mitigation strategies cannot be determined. Demographic

trends are also part of the equation; for example, 24 hours warning of a hurricane landfall might be adequate for a coastal city today, but not in ten or twenty years, after both population growth and the increasing occupation of coastal regions (as is happening in the United States).

Natural hazards researchers looking for inspiration about collaboration with social scientists can find it in the field of climate change, where discussions of limiting carbon dioxide emissions have spawned much interdisciplinary activity. Perhaps an even better model, being both more wide-ranging across disciplines and more immediate in its applications, is the effect of El Niño, an episodic warm-water current in the South Atlantic, on regional climate, fisheries and agriculture. Forecasts of El Niño are now routinely made, and attention has turned to how these can be used by governments and local communities to mitigate the effects of adverse climate variation on the regional economy. The International Research Institute for climate prediction, established last year by the US National Oceanographic and Atmospheric Administration, aims to “help translate scientific discoveries into beneficial social policies and strategies”. Its progress should be worth watching.

But one of the biggest obstacles to effective cooperation between natural scientists, social scientists and policy-makers is the difficulty in raising funds for cross-disciplinary activities. A 1995 report from the US National Academy of Sciences, called *Science, Policy and the Coast: Improving Decisionmaking*, concluded that more resources were needed for the application of coastal science to policy-making. Natural scientists may worry that the social scientists are just more mouths to be fed from the same shrinking pie. But these days, funding is tied increasingly strongly to a relevance to social needs. What better way to demonstrate such relevance than to show how scientific results can be translated into policy or practice? □

Coming down to Earth

Last week's UN meeting in New York was a salutary lesson in the realities of environmental politics.

If Rio was the party, then New York was the settling of the bill. Five years ago, 30,000 official and unofficial delegates from across the world gathered in Rio de Janeiro at the first Earth Summit to give a heady welcome to a new era of environmental concern. Last week, about one-tenth of that number gathered in New York to learn at first hand that deep-rooted social change is seldom achieved overnight.

There is some justification in their disappointment at the meeting itself. Despite an over-ambitious list of goals, there were at least some items on the agenda on which more progress, in different conditions, might have been achieved. One, for example, was the proposal for an international treaty on preserving the world's dwindling forests. Equally contentious, but still worthy of more support than it received, was Europe's suggestion for increased taxes on aviation fuel. In contrast, there was never any serious prospect that US President Bill Clinton would show his hand on the question of mandatory lim-

its on carbon emissions so far in advance of the Kyoto meeting in December, at which this issue is due to be thrashed out.

Perhaps the most important lesson of the meeting was that it underlined that a sound environmental policy requires two essential ingredients. The first is a solid scientific consensus on the nature of the problem; the second is respect for the democratic process, without which any environmental agreement lacks political authority.

The reality of environmental politics was to be heard less in the conference hall than in the voices of those New York inhabitants interviewed on the streets outside, determined to keep their air-conditioned, gasoline-guzzling way of life at all costs. Similar individuals can be found on the streets of Rome or Sydney, and fainter echoes in Bangkok or Beijing. If effective action is to be taken to protect the world from global warming, it is these people who must be convinced of the need. □



Molecular biologists in Asia and Pacific plan research network

[TOKYO] Life scientists from more than 20 institutions in the Asia-Pacific region met in Tokyo last Saturday (28 June) to discuss creating a network of researchers and institutions to promote collaboration in molecular biology and biomedical research.

Supporters of what has tentatively been called the International Molecular Biology Network (IMBN) for Asia and the Pacific Rim hope that it may eventually play a similar role to that of the European Molecular Biology Organization (EMBO).

The network is the brainchild of researchers at the Institute for Molecular Biology and Genetics at Seoul National University in South Korea, and the Institute of Medical Science of Tokyo University, where the meeting was held. Participants also included scientists from China, Taiwan, Hong Kong, Singapore, Malaysia, Australia, and New Zealand.

Sunyoung Kim of the Seoul institute, one of the organizers of the meeting, says that he hopes the network will awaken interest in molecular biology in parts of Asia where the subject is still in the early stages of development.

Cooperation between institutions in Asia is at present limited for complex historical, political and geographical reasons, and links

tend to be stronger with the United States, where much of the innovation in the field occurs. But collaboration between Kim's institute and Tokyo University over the past four years has convinced researchers at both institutions of the value of intraregional cooperation.

Other participants at the meeting were equally enthusiastic about improving links through regional postdoctoral fellowship schemes, conferences, workshops, short-term visits and practical courses. There was even talk of establishing a regional laboratory or laboratories, similar to the European Molecular Biology Laboratory in Heidelberg, Germany, which was set up by EMBO.

But there was no consensus on the form the network should take, or on its initial goals. For example, a draft document circulated at the meeting proposed two levels of membership, institutional and individual. But participants from China expressed reservations about a system based on institutions, without broad national and government backing.

The Chinese scientists proposed instead that branches be set up in each country to draw support from all institutions and the government. "If only one or two institutes from China join, others will quickly lose

interest," said Wei-Feng Chen of the department of immunology at Beijing Medical University.

There were also divergent views on individual membership. The draft proposed that any interested individual could join. But Frank Gannon, the executive director of EMBO who acted as an adviser to the meeting, pointed out that some degree of selection was essential to achieve scientific 'excellence'.

By the end of the meeting the participants were leaning in favour of a system of selective individual membership, where each participant, perhaps aided by neutral advice from EMBO, will nominate between five and ten outstanding scientists from their country, but allowances will be made in developing countries for the definition of 'outstanding'.

Also unresolved is the source of funding. The meeting was attended by senior officials from Japan's Ministry of Education, Science, Sports and Culture, the Science and Technology Agency and the Ministry of Foreign Affairs, as well as officials of the Korea Research Foundation, which is affiliated to South Korea's education ministry. They voiced support for the concept of IMBN but did not discuss where funds might come from, and could only point to existing or soon-to-be launched programmes that indirectly support the goals of the network.

Some participants expressed fears that, if only one or two countries provide funds for the network, they might dominate its activities. But Ken-ichi Arai, chairman of the department of molecular and developmental biology at Tokyo University's Institute of Medical Science, and an organizer of the meeting, believes that once the network is more clearly defined it will be possible to approach the Asia-Pacific Economic Cooperation (APEC) forum for support.

APEC, whose members include all the Asia-Pacific countries represented at the IMBN meeting, recently initiated meetings of its science ministers to develop intra-regional cooperation in science and technology (see *Nature* 384, 200; 1996), and the new network is closely aligned with APEC's goals.

One proposal discussed at the meeting was the setting-up of a website to link institutions, as an easy first step to develop the network. The meeting closed with a decision to set up a task force with seven members from Japan, Korea, Taiwan, China, Hong Kong, Singapore and New Zealand to seek 'seed' money of about \$2 million from various sources in order to run the network for the first one to two years.

David Swinbanks

Promising signs for NSF budget increase

[WASHINGTON] A key subcommittee of the US House of Representatives has proposed a funding increase for the National Science Foundation (NSF) of almost 7 per cent, suggesting that the agency could receive a substantial cash boost next year.

The foundation would get an extra \$90 million to rebuild its research station at the South Pole, in the proposal agreed by the appropriations subcommittee chaired by Jerry Lewis (Republican, California).

The subcommittee's proposal has still to be passed by the full Appropriations Committee and the entire House, and then reconciled with a yet-to-be-determined Senate proposal. But the decision has already raised the hopes of scientific societies in Washington, which had called for a 7 per cent increase for the NSF earlier this year (see *Nature* 386, 7; 1997).

Officials of these organizations say the decision reflects a growing acceptance in the Congress that politicians will get credit for investing money in science — and indicates how the scientific community can have an impact on the process. "When people raise



Lewis: seeks double Clinton increase.

issues, the people in Congress do pay attention," says Mike Lubell, head of public affairs at the American Physical Society.

Under a bill passed by Lewis's subcommittee on 17 June, the NSF would get \$3.5 billion in the 1998 financial year which starts on 1 October next — 6.6 per cent more than last year, and just over twice the increase requested by the Clinton administration in February (see *Nature* 385, 569; 1997).

Most of the extra money would go to the Antarctic programme. But money for research in universities would also be increased by 4.3 per cent. The equivalent Senate subcommittee, which has been given a smaller allocation of money than Lewis's, is expected to offer less to the NSF. But advocates of the agency say they hope it will propose an increase greater than the anticipated rate of inflation. **Colin Macilwain**

Quake panel admits prediction is 'difficult'

[TOKYO] The government committee responsible for overseeing Japan's earthquake prediction programme has approved a report by seismologists that clearly admits for the first time that earthquake prediction is difficult.

In the particular case of the expected Tokai earthquake near Tokyo, on which much of Japan's prediction efforts have focused, the review says that prediction will be possible only in a complex set of circumstances that are unlikely to be met.

The report is being widely interpreted in the Japanese media as admitting that earthquake prediction is virtually impossible. But government agencies involved in the research, as well as leading politicians, seem determined that the prediction programme should continue. The programme employs hundreds of researchers on an annual budget of more than ¥20 billion (US\$185 million) and has run for 30 years.

The report was put together by various subcommittees of the Geodetic Council, an advisory body to the Ministry of Education, Science, Sports and Culture made up of astronomers and Earth and space scientists.

These subcommittees were composed primarily of seismologists working on the programme. An external panel of scientists not involved in the research also produced comments on a draft of the report.

The report, which was approved by the

full council last week, concludes that predicting the time, location and magnitude of earthquakes is, in general, *koonan* ("difficult"). The word has connotations of being almost impossible. But its meaning is ambiguous — allowing government agencies and researchers to justify continued research on prediction.

Huge investments have been made in setting up dense observation networks in the Tokai region around the Izu peninsula, as researchers have predicted that a major earthquake can be expected in the near future. A committee of six 'wise men' can be called in by the meteorological agency at any time to examine seismic data that might indicate that an earthquake is imminent, and advise the prime minister to issue a warning.

The leaders of the programme have always maintained that prediction of the Tokai earthquake is possible, and this has provided much of the justification for their work. Their confidence was based in part on the belief that the Tonankai earthquake of magnitude 7.9 that occurred in the same region in 1944 showed clear precursors that could have acted as a warning.

But the quality of observation data in 1944 was limited, and much of the 30-year programme has been spent collecting empirical data in the hope of finding clear-cut precursors of earthquakes.

The new review for the first time places

severe limitations on the possibility of predicting the Tokai earthquake, however, by stating that, if prediction is to be possible, crustal motions preceding the earthquake must be of the same type and follow the same time dependence and amplitude as the Tonankai earthquake.

In all other circumstances, prediction of the Tokai earthquake will be "difficult", the review says, adding that no estimate can be given of the likelihood that the required conditions will be met.

For the first time in the history of the programme, the report accepts that there is a "huge gap" between the public's perception of researchers' ability to predict earthquakes and the present level of the science. It calls for the "difficulties" of earthquake prediction to be made "widely known" to the public, which has become disillusioned with the science since the Kobe earthquake of 1995 which caught the public completely by surprise. It also calls for greater efforts in disaster mitigation.

Despite such admissions, the government agencies backing the programme have expressed their determination to carry on with the research. The chief cabinet secretary, Seiroku Kajiyama, has said: "It is a matter of course that the government should continue efforts [to predict earthquakes] if there is a possibility of earthquake prediction in the future".

David Swinbanks

German law could boost prospects for organ transplants

[MUNICH] The number of organ donors in Germany may soon increase following a decision last week by the Bundestag, the lower house of the federal parliament, to approve a controversial bill regulating organ transplants.

The bill, which has had a five-year struggle, is expected to be approved by the upper house, the Bundesrat, in the next few weeks. It gives legal status to a practice that has in fact been carried out in Germany for many years.

The long debate about the law is due to Germany's extreme sensitivity on any bioethical issue, a reaction to the country's Nazi past. Opponents fear that organs could be removed from people who are still alive.

The German Chamber of Physicians first drew up a protocol for organ transplants in 1982. That, in line with other European countries, defined death as the moment at which the brain ceases to function.

But efforts to formalize this in law, begun in the early 1990s, led to an emotional debate in which critics from all political parties argued that death should be defined as the moment at which the heart also stops

beating. As a result of the controversy, the number of organ donors fell by 10 per cent.

The new law does little more than confirm present organ transplantation practice, which is slightly more conservative than elsewhere in Europe. Organs may be removed from a donor at the moment of total brain death, rather than after brain-stem death, which many countries accept as the first irreversible event of death.

Brain death must be confirmed by two independent doctors. Previous 'informed consent' from the donor, through a standard donor-card system, or through verbal or written consent in hospital, is required. In its absence, a relative or partner may give consent, but may not contradict any known wishes of the donor. Only close relatives of a patient, or a spouse, may donate kidneys as live donors. The sale of organs is banned.

Organ allocation will still be the task of Eurotransplant, the Netherlands-based office, which uses a single database to allocate organs from its five member countries — Germany, the Netherlands, Austria, Belgium and Luxembourg.

Bernard Cohen, director of Eurotransplant, which allocates 5,000 organs a year but which has a waiting list of 15,000, says the greater public confidence resulting from the new law might increase organ donations slightly, or at least lower the refusal rate.

But he says he would not expect a large increase in the number of donors, because the requirement for informed consent has a restricting effect. The two members of Eurotransplant with the lowest per capita rate of organ donation are the Netherlands and Germany, which both require informed consent, he says.

According to Ludger Honnefelder, director of the Bonn Institute for Science and Ethics, the law has been carefully drafted to avoid directly equating death with brain death to side-step the problem of what 'alive' really means.

Reflecting continued memory of Nazi practices, the proposed law states in two separate paragraphs that organs may not be removed until death is observed and that organs may not be removed before total brain death is registered.

Alison Abbott

US air pollution rules stir up a stink about research

[WASHINGTON] The decision last week by President Bill Clinton to approve controversial new US regulations on air pollution — despite opposition from industry and some of his own economic advisers — is expected to provoke a congressional challenge when the rules are published later this month.

Ironically, it may also give scientists something they have been requesting for years: more government spending on research to reduce the scientific uncertainties surrounding the effects on health of ozone and small airborne particulate matter. That uncertainty helped to fuel months of rancorous debate, with each side accusing the other of “junk science”, and researchers being caught in the political crossfire.

Clinton endorsed regulations only slightly looser than those proposed by the Environmental Protection Agency (EPA) last November (see *Nature* 384, 392; 1996). The 24-hour allowable standard for fine particulate matter (2.5 microns and smaller) will be set at 65 instead of 50 micrograms per cubic metre. Ozone polluters will be charged after the

fourth time they exceed acceptable levels rather than the third. And municipalities will not have to comply with the new rules for at least seven or eight years.

The EPA administrator, Carol Browner, said that it will take up to five years just to set up a monitoring network for fine particles, which are not regulated at present. The delay, she said, “gives us a chance to make sure that all the science is made use of and is thoroughly reviewed”.

Some scientists, frustrated at the perennial lack of funding for research in this area, believe that issuing a regulation now, even with many gaps in our understanding of the health effects of the pollutants, may be the only way to stimulate greater investment. Joe Mauderly of the Lovelace Respiratory Research Institute in Albuquerque, New Mexico, who chairs EPA's Clean Air Scientific Advisory Committee (CASAC), told Congress last March: “I think it's unfortunate, but probably reality, that information [on particulates] will not be collected unless a standard is promulgated.”



Hazy view: scientists lack funding for research on smog (above) and other pollution problems.

The message may finally be getting through. On the same day that Clinton endorsed the regulations, a House of Representatives appropriations subcommittee added \$40 million for ozone and particulates research, on top of the \$45 million the administration had already requested for 1998. But not all the increase would necessarily go to EPA. The subcommittee called for the National Institute of Environmental Health Sciences, part of the National Institutes of Health, to become involved in distributing the research money.

Others in Congress say they will fight the new regulations. One option would be for Congress to test its new authority to veto individual regulations. Another would be simply to hold up EPA funds for implementing the rules through the appropriations process.

Some scientists are still smarting from political tactics used during the debate. The House Commerce Committee chairman, Thomas Bliley (Republican, Virginia), who opposes tighter regulations, has been putting pressure on Douglas Dockery of the Harvard School of Public Health and his colleagues to release raw data from a study showing increased sickness and death from exposure to particulate matter.

When EPA joined in the request last spring, the Harvard team argued that releasing questionnaires containing personal data about respondents would violate confidentiality agreements. But Harvard finally agreed to an independent reanalysis of the Harvard study, with controlled access to the questionnaires. This will be done by a panel of scientists convened by the Health Effects Institute in Cambridge, Massachusetts, which is funded jointly by EPA and industry.

The institute sponsored a workshop last week to determine how the reanalysis of this and two other studies on the health effects of particulates will be done. But that has scarcely toned down the rhetoric. Paul Beckner, president of the pro-industry Citizens for a Sound Economy, said last week: “In the interest of truth, we call on the administration to make public the secret data that the EPA has long claimed justifies its proposal.” **Tony Reichhardt**

President of UN summit ‘sobered’ by outcome

[WASHINGTON] A special session of the United Nations general assembly closed in New York last week with few signs of progress towards the goals agreed at the 1992 Earth Summit in Rio de Janeiro, and a growing sense that nations are reluctant to pay anything more than lip-service to them.

The session was marked by both its high level of attendance — the leaders of most of the world's leading economies delivered speeches — and low expectations, with little progress anticipated or achieved. But even these expectations were dashed when the parties failed to agree on a broad-ranging “political statement” intended to point the way towards meeting the Rio goals.

The meeting ended instead with prolonged negotiations to produce a more lengthy and detailed programme for the further implementation of Agenda 21,

the document agreed at Rio.

The president of the session, Ismail Razali of Malaysia, admitted that “the overall results” of the special session were “sobering”. Environmentalists went further. “The political statement, which was supposed to be a wake-up call to the world, has been abandoned,” says Bill Hare of Greenpeace, the environmental group. “This is verging on a disaster. It isn't clear what credibility this entire process has left.”

Officials at the UN pointed out that some progress was made, with agreement for the first time to phase out lead in petrol worldwide, for example, and plans to establish mechanisms for global action on freshwater and forestry conservation.

But this headway was overshadowed by public deadlock on the issue of global warming. Pressure from Europe and the small island states failed to extract

specific promises from Japan or the United States on cuts in greenhouse gas emissions.

Last Thursday (26 June), US President Bill Clinton told the meeting that he would take action “to convince the American people and the Congress that the climate change problem is real and imminent”. He added that he would bring to December's meeting at Kyoto, Japan, where countries hope to negotiate a treaty on greenhouse gas emissions, “a strong American commitment to realistic and binding limits that will significantly reduce our emissions of greenhouse gases”.

The US administration has yet to decide what kind of limit it will pursue at Kyoto. It may seek anything from 10 per cent below to 10 per cent above 1990 emission levels by 2010, according to one environmentalist.

Colin Macilwain

Cloning for research 'should be allowed'

[WASHINGTON] The British leader of the research team that cloned Dolly the sheep has strongly endorsed US moves to outlaw the cloning of human beings. But he said that he has no ethical objection to the use of his technique to create for research purposes human embryos that are not implanted.

Ian Wilmut of the Roslin Institute near Edinburgh said it is "entirely appropriate" to decide that cloning human beings is "not socially acceptable and for a law to be passed". Wilmut was speaking last week at a seminar sponsored by the American Association for the Advancement of Science.

His comments came three days after leaders of the world's eight major industrial nations issued a communiqué after their summit meeting in Denver, Colorado, stating agreement on "the need for appropriate domestic measures and close international cooperation to prohibit the use of somatic cell nuclear transfer to create a child".

The heads of government who agreed to the statement came from Canada, France, Germany, Italy, Japan, Russia, the United Kingdom and the United States.

President Bill Clinton has already sent to Congress a bill that would penalize anyone attempting "to create a human being using somatic cell nuclear transfer cloning" (see

Nature 387, 644 & 748; 1997). But anti-abortion advocates have complained that the proposed law — as reflected in the Denver communiqué — would not penalize the use of the cloning technology to create human embryos where such work stops short of implantation.

Wilmut said he would have no ethical difficulties with such research. But he would have a practical objection: "There are a very limited number of oocytes available for research with human embryos. For the foreseeable future this technology would be much more appropriately developed in a laboratory animal."

His comments coincided with reports that many animals pregnant with clones in research laboratories in the United States and Europe are miscarrying, and that some of the surviving fetuses show evidence of subtle genetic abnormalities. Others are growing abnormally large in the womb.

Other speakers at last week's seminar were less sanguine than Wilmut about the prospect of a law banning cloning. Maxine Singer, president of the Carnegie Foundation, warned of a "slippery slope" that could eventually lead to Congress banning other kinds of research. "To make a precedent like this, to have national legislation that would

govern what people can do in labs, would be a very, very big step," said Singer.

Gillian Woollett, assistant vice-president of biologics and biotechnology at Pharmaceutical Research and Manufacturers of America, said one concern is that the Clinton bill and the report from the National Bioethics Advisory Commission on which it is based contain between them "three different definitions of a somatic cell".

The bill defines a somatic cell as "any cell of the body other than germ cells (egg or sperm)". The bioethics commission's 110-page report, published on 9 June, defines a somatic cell in its glossary as "any cell of an embryo, fetus, child or adult not destined to become a sperm or egg cell". And on the first page of the report, the commission defines a somatic cell as "any cell of the embryo, fetus, child or adult which contains a full complement of two sets of chromosomes; in contrast with a germ cell, i.e., an egg or a sperm, which contains only one set of chromosomes".

One influential member of the biotechnology industry argued that federal action to prevent the cloning of human beings would be preferable to a patchwork of state laws. "If there is not a sufficient national response you will find very, very unfavourable, awkward and in some cases very misinformed legislation cropping up in the states," said Carl Feldbaum, president of the Biotechnology Industry Organization.

He cited a bill introduced in the Florida state legislature, which would have inadvertently banned the 'cloning' of human DNA through *in vitro* replication.

But Feldbaum, a criminal lawyer, said that the Clinton bill still needs "enormous work," partly because of its lack of a declaration that it would pre-empt state law. He warned that the bill proposed "draconian" penalties and an indistinct intent clause that could deter legitimate research.

Jeff Smith, the executive director for policy at the White House's Office of Science and Technology Policy, said that the White House had received assurances from James Sensenbrenner (Republican, Wisconsin), chairman of the House Science Committee, that he would produce anti-cloning legislation. "Whether it will be an exact clone of the president's bill is not clear," Smith said. The purpose behind the Clinton bill, he said, was to frame the debate and "get the president's position out in front".

Meredith Wadman



GARY TRAMONTINA/AP

European biotech industry plans ethics panel

[PARIS] Europe's biotechnology industry is seeking to improve its image by setting up a committee of independent experts to advise it on ethical issues and to draft a code of conduct for its members.

The committee is the initiative of EuropaBio, an industry association representing about 600 biotechnology companies and 11 national biotechnology industry associations. Andrew Dickson, the secretary general of the organization, says that the industry felt it "needed not just to be acting responsibly but to be seen to be acting responsibly".

EuropaBio has not yet decided on the final form of the committee. It says it would provide a secretariat but that the eight to ten members of the committee would need to be — and to be seen to be — independent

of industry, with complete freedom in their activities. Dickson says EuropaBio is contemplating asking neutral bodies to help to nominate the members, who would be paid only expenses.

The ethics committee would be modelled on the European Commission's expert advisory group on biotechnology ethics, which is chaired by Noëlle Lenoir, a member of the French constitutional council.

EuropaBio last week issued a draft "core set of ethical values", which it has made available for public comment (website: <http://www.europa-bio.be>) and to which its member companies will be expected to adhere. This would be revised regularly to take into account scientific progress and changes in public perception.

Dickson says it is important for the industry to

take the temperature of "the boundaries which society believes are acceptable" and to act within these if it is to win public acceptance.

The current draft of the code of ethics rejects the use of cloning to reproduce human beings, arguing that this reflects the current consensus. But Jurgen Drews, chairman of EuropaBio and president of research at Hoffmann-La Roche, points out that this consensus may change.

The code also commits member companies not to work on genes of human sperm, eggs or germline cells. It proposes a moratorium on work on the genes of human embryos "until the medical, ethical, and societal issues that may arise from this kind of therapy have been publicly discussed, clarified and resolved and, if applicable, put into legislation".

Declan Butler

France puts big science under scrutiny

[PARIS] France is to re-evaluate its commitments to international 'big-science' projects and its own national research facilities.

Officials at the science ministry decline to give specific examples of which such projects might be targets for cuts. But the move may raise questions over the terms of France's participation in facilities such as the planned international space station and the European Laboratory for Particle Physics (CERN).

The new socialist government also plans a radical reform of its national research agencies, and the way they evaluate laboratories and grant applications. It will substantially increase the recruitment of scientists, and will create a system of short-term post-doctoral fellowships in a country where lifetime employment has been the rule.

Government sources say that Claude Allègre, the minister of national education, research and technology, is deeply sceptical of the value of many big-science facilities, believing they often escape proper evaluation.

According to one ministry official, the prowess of powerful scientific lobbies has often been the main factor in persuading politicians of the case for constructing bigger and better machines. Proposals for new national facilities will be considered only if the possibility of international funding has been "extensively" explored, he adds.

Ocean drilling

Big science is one area where the cash-strapped government feels it can make savings, says the official, adding that both national facilities and France's international commitments will be reviewed.

One project likely to come under close scrutiny is the international Ocean Drilling Program, whose value for money has already been questioned (see *Nature* 379, 193; 1996). Allègre, an Earth scientist by background, is also known to be sceptical of the usefulness of manned space flight, including the planned international space station.

The government's hands are tied to some extent by the fact that responsibility for space policy is shared with the president, Jacques Chirac. But the government intends to appoint a new director-general of the French space agency CNES, for which Allègre is responsible, to reorientate the agency's activities. Support for CNES has previously been motivated by national prestige. "CNES in the future will be driven far more by technical and scientific aspects," says the official.

Allègre last week gave his first press conference since becoming minister. He confirmed his intention to end staff cuts throughout the research and education systems, and to create "hundreds" of new posts for researchers, technicians and lecturers within the research organizations and

universities by as early as September. This would be followed by sustained high levels of recruitment to offset the scheduled massive wave of retirement over the next decade.

The science ministry has some margin to allow it to make such new investments. Lionel Jospin, the prime minister, has confirmed that research and education will be national priorities and that this will be one of the few ministries to receive increased funding in the September budget. But Vincent Courtillot, Allègre's principal adviser, says the increase will be "significant but not overwhelming".

Massive increases in public spending have been ruled out as a result of the government's obligation to meet the budgetary criteria for joining the single European currency, and some financial acrobatics will be needed. Allègre hopes to finance increases for his main goals, for example, by making cuts and economies elsewhere within his existing budget.

The possibilities for redistributing money in this way are, in principle, favoured by the unprecedentedly broad scope of Allègre's 'superministry', which has responsibility for virtually all aspects of education, science and technology.

Research organizations and universities are concerned, however, that the increase in recruitment could actually make matters worse unless it is accompanied by extra research funds, as salary costs already eat up most of the research agencies' budgets.

Salaries already account for 80 per cent of the budget of the Centre National de la Recherche Scientifique (CNRS), for example, and despite large staff cuts this proportion grew by 3 per cent last year, whereas the agency's total budget grew by just 1.5 per cent. An increase in recruitment would seriously exacerbate this situation, unless it were accompanied by a large increase in CNRS's overall funding, says Pierre Tambourin, director of the agency's life sciences division.

At the same time, the government plans to introduce a system of postdoctoral fellowships. It says this is needed to absorb the 10,000 to 15,000 unemployed postdoctoral researchers and to promote mobility, but that the ultimate aim is for postdoctoral researchers to find permanent jobs. To encourage recruitment in the private sector, it proposes increasing tax credits on research for small and medium-sized companies, while indexing these directly to specific recruitment targets.

Henri-Edouard Audier, a chemist at the Ecole Polytechnique and a member of the national bureau of the National Union of Scientific Researchers (SNCS), says that the unions accept the government's arguments



about the need for more flexibility as a provisional measure. But it does not want this to become a slippery slope towards the Anglo-Saxon system, with a large population of researchers on short-term contracts.

Shake-up for CNRS

The research organizations themselves face the prospect of radical shake-up. They have been asked to justify their activities, as well as to identify areas of duplication.

"We need to look at the entire research landscape from scratch," says Tambourin, arguing that a large institute for human pathologies could regroup the biomedical research agency INSERM and parts of CNRS, while a new agency dedicated to studies of the planet could regroup research on biodiversity, the environment and other areas.

Orstom, the agency for research in developing countries, seems almost certain to be abolished, for many now in power consider it to be an anachronism of colonialism, and as much an instrument of French foreign policy as a research agency. Its core activities are likely to be redistributed among the other agencies.

The unions have previously resisted the expected reforms in the allocation of jobs and money. But they are now making conciliatory noises. Audier concedes that the goals of the research organizations need to be redefined, and says this will be acceptable provided the changes are carried out in genuine consultation with all the parties involved.

Audier is optimistic that the national priority now being given to research has improved the climate for negotiating this and other reforms. But there has already been one major disagreement between the unions and the ministry. Allègre wants to reduce drastically the size of the research agency committees that evaluate laboratories and set scientific strategy, and in particular CNRS's powerful national committee, which consists of commissions covering 40 disciplines.

Declan Butler

Uncertainty over fate of Mir experiments

[WASHINGTON] Russian and US engineers and scientists were this week taking stock of the damaged Mir space station to see whether power can be restored to its laboratory modules and what experiments — if any — can still be performed. The future of the 11-year-old craft, which has lasted far longer than any other Russian space station, hangs in the balance.

The Mir sustained serious damage last week when an unmanned cargo ship crashed into its Spektr laboratory module during a test of a new manual docking system. The crash damaged one of the station's power-producing solar arrays and opened a small hole in the Spektr, which immediately began leaking its pressurized atmosphere.

That forced the two Mir cosmonauts and a US guest astronaut to disconnect power from three other arrays and seal the module off from the rest of the station. The equipment inside Spektr, roughly half of which was provided by the US space agency NASA, is now in a cold vacuum, and the experiments inside are thought to be lost.

But scientists working on the joint US-Russian research programme will not know for certain until they take an inventory of the equipment that was inside Spektr when the accident occurred. Some is portable and may

have been moved elsewhere on the station.

The 1,600 pounds of US-provided equipment inside Spektr includes a large freezer for storing blood samples, a centrifuge and equipment for cardiology investigations. The French national space agency, CNES, is also assessing the status of its equipment inside the module. Russian experiments in Spektr are concerned primarily with Earth observation and atmospheric studies.

Priroda, the laboratory module that contains US crystal growth experiments and other materials-science investigations, was not harmed. But these experiments require much power to run furnaces and other equipment. The now disconnected Spektr solar arrays provided about half the station's total electrical power. These would need to be reconnected to restore Mir's full research capabilities.

"The big unknown for [future] science [on Mir] is how the power shakes out," says Thomas Sullivan, a NASA scientist working on the joint research programme.

The current plan is for Mir's two cosmonauts to enter Spektr wearing spacesuits in mid-July and try to reconnect the power cables. The crew has never trained for such a job, and no one is sure if the repair is possible.

NASA officials and others were quick to

point out that last week's accident had nothing to do with Mir's age. But the Russian space programme is clearly strained to breaking point. It also badly needs the revenue that comes from renting space on Mir.

France has signed a deal for a reported \$40 million to send an astronaut to the station in August, and again in 1999. Germany has also paid to use Mir, and the United States is spending \$472 million for seven long-term stays by NASA astronauts — two of which have yet to happen.

Last year, 25 per cent of the Russian Space Agency's funding came from renting out Mir, according to an analysis by ANSER Corporation of Washington DC. "Without that [revenue] it will be tough going," says Stephen Hopkins of ANSER.

Many US space officials worry that money spent patching up Mir will make it harder for Russia to meet its financial obligations to the international space station scheduled to begin construction next year.

Meanwhile, James Sensenbrenner (Republican, Wisconsin), chairman of the House Science Committee, has called on NASA administrator Daniel Goldin not to send another US astronaut to Mir until NASA can certify that the Russian station meets or exceeds US safety standards. **Tony Reichhardt**

Australian science review defends diversity and overlaps

[SYDNEY] Australia's chief scientist, John Stocker, who was asked by the government in February to carry out a review of the organization of Australian science, has recommended a largely uncontroversial streamlining of advisory mechanisms.

But in his report, published this week, Stocker argues against identifying "any gaps or overlaps", as he had been asked to do by the government. Gaps in research are "unavoidable" in a country the size of Australia, he says, while overlaps are "necessary and desirable".

The country's largest research agency, the Commonwealth Scientific and Industrial Research Organization (CSIRO), had been pushing to absorb the Australian Institute of Marine Science (AIMS) and possibly the Australian Nuclear Science and Technology Organization (ANSTO). AIMS is responsible for research in tropical waters, especially around the Great Barrier Reef and coral reefs off northwest Australia.

But Stocker comes down strongly in favour of diversity and plurality in publicly supported research, rejecting any concentration of research institutions within the portfolio of the Science and Technology Minister, Peter McGauran, who commissioned the review.



Reichtel: saw marine science prioritized.

Stocker supports the efforts of the smaller agencies to retain their specialized identities, and their claims that amalgamation would bring added costs and demands to hand over some of their resources. But he recommends that the three bodies should work out a common strategic plan.

Senior staff at AIMS are relieved that Stocker's proposals appear to reduce the threat to their independence. Russell Reichelt, director of AIMS, says he is "gratified" that marine science, which Stocker says is underfunded, is seen as a "high priority area".

Helen Garnett, executive director of ANSTO, argued in a statement that "the current pluralist system will continue to deliver best benefit". But CSIRO's chief executive, Malcolm McIntosh, repeated his organization's view that "it may be desirable for Australia to have fewer, larger and better-equipped institutions".

Stocker's report implicitly criticizes the cost-cutting strategy pursued by the Coalition government since last year's

election. He says that incentives are necessary to boost business research and development, a view backed by numerous submissions. "My personal view is that it was not a good idea to reduce the tax break for industry from 150 to 125 per cent."

The main message of his report is that Australia urgently needs to develop policies for science and industry and to identify priorities if it is not to be left behind in the global technological market. He suggests that the Prime Minister's Science and Engineering Council, currently little more than a discussion group, should be given the authority to launch initiatives and to have their implementation overseen by a cabinet committee.

While declining to comment on whether the science minister should become a full member of the cabinet, Stocker argues that the minister's profile would be raised through membership of such a committee.

The impact of Stocker's proposals will depend on the reaction of the Prime Minister, John Howard, and on whether the cabinet is prepared to allow McGauran to take on a more powerful role. McGauran issued a statement saying that he needs to consult other ministers before the government issues a response. **Peter Pockley**

Asian tiger claws its way up the ratings

Philip Campbell

Taiwan has ambitious plans to achieve international stature in science, and is investing accordingly. It is pursuing long-term national goals for technology development and to boost support for its best scientists.

Visit top officials of Taiwan's National Science Council (NSC) and you will meet a group of determined individuals. Closely tracking the strengths and weaknesses in their country's scientific and technological development, they are dedicated to keeping Taiwan moving up the world scientific rankings, and to diminishing its technological dependence on others. And they have money to do it.

The NSC is responsible for research evaluation, development of science policy and funding of basic research in universities, and plays a key role not only in Taiwan's science, but also in its high-technology industrial development. It has several advantages. Despite Taiwan's political isolation, the country is sufficiently robust to have developed significant economic clout. With a population of 21 million, and the 'kickstart' of wealth and technocrats brought from mainland China when Chiang Kai-Shek and his followers fled from Mao Tse-tung's revolution in 1949, the country has developed an economy currently growing at a healthy 6 per cent a year.

Its science statistics also tell an encouraging story. Taiwan achieves respectable levels in rankings of influential patents (seventh in number of patents approved by US Patent and Trademark Office) and scientific papers (19th in number of papers in the *Science Citation Index*). On a regional basis, NSC officials have discovered that Taiwanese researchers come top (against Japan, South Korea and mainland China) when ranked by

the proportions of papers published in the six leading scientific journals in each of physics, chemistry and mathematics.

The NSC and the Academia Sinica (Taiwan's national academy of sciences) carry the prime responsibility for developing this growth further, fostering collaboration between all elements of the research base wherever possible.

Heading the NSC is Chao-Shiuan Liu, one-time chemist and past minister of telecommunications. His position gives him a central role in implementing the 'National Strategic Plan' for science and technology, which was endorsed at a national conference last September attended by leading academics, industrialists and administrators.

One key goal of that plan is to boost the country's overall expenditure on research and development from its current level of 1.8 per cent of gross domestic product to 2.5 per cent — the same as South Korea — by the year 2000. That, as Liu acknowledges, is a tall order, although the trend is in the right direction. Last month, it was announced that the government's R&D budget would be increased by 10 per cent in the next fiscal year, from 1 July (see graph overleaf). But it is hard to imagine how this figure can grow at an average annual rate of 15 per cent over the next few years, as the plan's target demands.

Meanwhile industry, also urged to increase its spending on R&D, lacks the manpower to do so. The national plan includes means by which that skill shortage can be overcome.

"My immediate priority is to loosen constraints," says Liu. Chief among these are obstacles to direct investment in industry — Liu argues that "new tax incentives are required" — and rigidity in employment laws. Greater flexibility in government procurement is also required. "We need to consider quality as well as price," says Liu. To achieve these ends, a series of laws will be presented to the Congress within the next few months.

Also important to the plan is cooperation between the NSC and the Academia Sinica, many of whose 124 institutes and other centres will play a key role. The academy's president Yuan Lee — joint winner of the Nobel Prize for Chemistry in 1986 for his work on the dynamics of chemical elementary processes — is said to have a closer relationship with Liu, who took over at the NSC a year ago, than with his predecessor.

Lee is committed to strengthening the relevance of Taiwan's research institutions and university laboratories to the country's development. "We have many people who have returned from the United States with policy experience who are beginning to influence government thinking," he says.

People

Although funding and technology targets are important, Taiwan's leaders are giving equal priority to developing hotspots of international excellence within the science community through a new 'Frontiers Sciences' programme.

"We asked ourselves 'who is there of world-class excellence in the life sciences in Taiwan?' and came up with about 20 names," says Steve Hsieh, vice-chairman of the NSC. The intention now is to select 10 of these, remove from them the burden of annual grant renewals (a frequency that Hsieh acknowledges is too high), and give them significant resources for up to five years. "We'll give them three or four times the average level of support," says Hsieh.

As elsewhere, researchers in Taiwan are heavily burdened by administrative demands. "These researchers have to be free of bureaucratic inflexibilities," says Hsieh. "They should be able to hire PhDs, travel to conferences and so on without having to apply for additional resources."

In establishing this liberal approach, Hsieh has had to overcome some resistance. "People worried that this would lead to abuse. That is a cultural obstacle — our society tends to look on the bad side. But we cannot make everything corruption-proof. Researchers are the élite of our population; we should trust them."

Younger scientists also need nurturing, says Hsieh. To that end, the NSC intends to introduce a parallel scheme, with five-year awards, for promising researchers. "We will cultivate the best of them, and then expect



All dressed up and somewhere to go: the president's palace in Taipei, capital of a country which is heading for world stature in science and engineering.

that they will move on into the Frontiers Sciences programme."

A key factor in Taiwan's scientific vigour is the fact that over the last two decades, at every level, excellent Chinese expatriate scientists — including Lee, Liu, the heads of many of the academy institutes, and professors and staff at the universities — have been attracted back from the United States. But that needs to change, says Wei-Jao Chen, president of Taiwan National University in Taipei. "We need to open up the culture by attracting people from Europe, for example."

Moreover, there are signs that Taiwan is now developing its own sources of excellence, as more and more vacancies are filled by Taiwanese. Hsieh is cautious about that. "This could have the disadvantage of a lack of internationalism in their viewpoint, so young scientists should still be encouraged to spend significant time overseas." Where? There is no doubt as to tradition and preference: every day, five flights depart from Taipei airport bound for California.

Links with researchers in the United States and use of the Internet both help to overcome the isolation in Taiwan. But, for those who have returned, it is hard to maintain standards of excellence. "Despite the good communications, we feel out of the loop," says one laboratory head.

"It's hard to define why. But people returning from the United States often have all the makings of an assistant professor over there. Given the limitations here currently, after 10 years here, they couldn't, for example, be full professors in the United States." Enhanced funding may help, and, to keep standards up, academy institutes organize annual reviews by international groups of scientists.

Worries about isolation may also feed another problem that occasionally under-

mines Taiwan's progress: competition between researchers keen to make their individual reputations. "People are reluctant to cooperate with other groups in the laboratory," says one academy institute director ruefully. "The result undermines us."

"For example, in a joint project on protein structure and function, the impact was weakened because the collaborators split the results into two publications to maximize individual credit."

Lee points to the need for strong leadership to overcome such problems: "We need senior people working together, which in turn requires a respected leader to drive it forward."

Integration

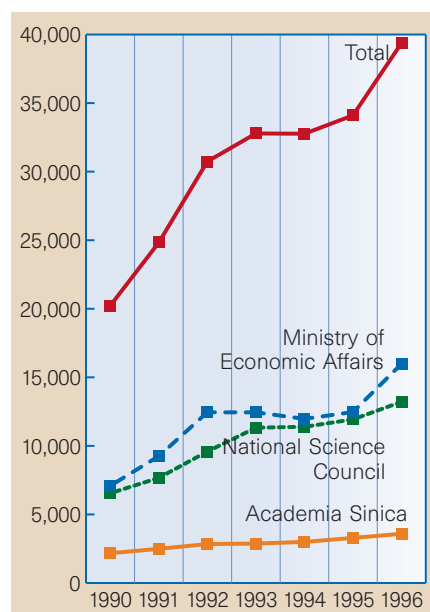
While support for individual researchers can be expected to stimulate excellence at the micro level, macro objectives also figure prominently in Taiwan's national plan, with the intention of strengthening interactions between science and industry in target sectors. That, it is hoped, will boost indigenous technological innovation and the flow of skills into industry.

The industrial base is too heavily dominated by small companies to achieve rapid innovation even with state encouragement, says Liu. To overcome that, the national plan identifies four 'national projects', in telecommunications, agricultural biotechnology, health biotechnology, and natural-disaster mitigation. The projects' management teams have been put together in recent weeks, and they will develop plans and budgets over the next few months.

In telecommunications, says Liu, the challenge is to reduce dependency on imports for technology development. After the deregulation of the industry, many small and medium-size companies have developed, supplying equipment, systems and operations.

"There is plenty of scope for rapid growth for years to come," says Liu. "When we started up this industry, we imported skills from the US. Now we have to breed our own, and one major goal of the project is to turn out people with appropriate postgraduate qualifications three or four years from now and to stimulate technological creativity. As in all the national projects, the key goal is to integrate our funding."

The history of biotechnology in Taiwan has been much less vigorous. Since 1990, government task forces have been set up to boost pharmaceuticals, livestock vaccines and biopesticides, culminating in the establishment of the National Health Research Institute in Taipei and an agricultural biotechnology zone in the science-based industrial park in Tainan. These developments and the immigration of Chinese biologists from the United States gives Taiwan good scope to boost its efforts in biotechnol-



Taiwan government spending on science and technology (in million NT\$). In 1994 (the latest year of available statistics), government spending accounted for about 48 per cent of the national total. Of that total, basic research accounted for 15 per cent, applied research 35 per cent and experimental development 50 per cent.

ogy, says Liu, and to that end national projects are being developed.

Biologists appear cautiously optimistic. "We have no strong pharmaceutical company and some small drug companies," says Ken Wu, head of the academy's National Health Research Institute. "But, with a new academy institute in agricultural biotechnology and the growth of science parks, things are shifting in the right direction. Government needs to make its top priority in decreasing regulation and improved coordination. We need coherence between academy and industry. And we need to achieve that within the next few years while the opportunity is there."

Another senior biologist is more blunt. "Coherence was a problem partly because the previous head of the NSC and the head of the academy didn't get on. Now things are much better. In the last two years people have tried to rush in with money to help set up companies, but we need better advice — consultants from the United States, perhaps."

"For example, the way in which intellectual property rights and conflicts of interest work are well understood in electronics here, but not in biology. Right now we are allocating funds for biotech, but I think it will take another two years to really get it sorted out."

Others share the view that dramatic changes cannot be expected overnight. But, in the longer term, Taiwan's scientists and policymakers appear confident that their country will achieve world stature in research and in innovation — and are happy to wait. Says Hsieh: "Give us ten years!" □



High tech, high hopes: the production of colour televisions is one of Taiwan's success stories, boosted by a desire to reduce technological dependence on other countries.

Venter opens access to sequence data on deadly pathogens

[WASHINGTON] Shortly after publicly splitting up with its former commercial partner, the Institute for Genomic Research last week released DNA sequences from 11 micro-organisms, including the pathogens that cause tuberculosis, cholera and the deadliest form of malaria.

Craig Venter, president and director of the non-profitmaking institute based in Rockville, Maryland, calls the decision to allow other researchers open access to sequences containing more than 40 million base pairs the largest single release of genomic information. He predicts that it will have a "tremendous impact on scientific research worldwide".

He says the move symbolizes the freedom the institute is now able to exercise after its split from Human Genome Sciences (HGS), the Rockville-based drug developer that had been its commercial partner since 1992. William Haseltine, chairman and chief executive officer of HGS, says his company is "delighted" with the "amicable" split which will have no effect on HGS, which has a substantial sequencing capacity of its own.

UK universities told to boost infrastructure

[LONDON] Britain's universities might consider reducing the amount of public money they spend directly on research, and spending more on maintaining buildings, facilities and research equipment, according to a report from the Parliamentary Office of Science and Technology.

The report says there is an annual shortfall of £670 million (US\$1.1 billion) between the cost of research and payments to universities from various sources. This is at present met by diverting money away from infrastructure spending, but the report suggests that the gap might be bridged if universities charged more for their research.

French academy split on cloning policy

[PARIS] A statement by the French Academy of Sciences condemning the cloning of human beings as "ethically unacceptable" has been challenged by several of its members. The academy says some members had opposed the statement or expressed reservations. They argued that it was not the role of the academy to take "ontological positions" and that, if a cloning ban was justified, it should be provisional.

The academy based its position on the argument that cloning would attack human dignity and autonomy, and open the way to

eugenics. But it expressed scepticism about the practicality of cloning, (see page 6).

Researcher suspended in fraud investigations

[MUNICH] Friedhelm Herrmann, a leading researcher at the University of Ulm in Baden-Württemberg, has been suspended from his post by the Baden-Württemberg ministry of science during investigations into allegations of large-scale scientific fraud. The fraud scandal, involving alleged fabrication of data in up to 30 scientific papers in molecular medicine, came to light earlier this year (see *Nature* 387, 442 & 750; 1997).

Herrmann denies charges of being implicated in the fraud. But his former colleague, Marion Brach, has admitted forging data in at least four papers, and claims that this was done with Herrmann's knowledge. Baden-Württemberg's science minister, Klaus von Trotha, decided to suspend Herrmann after an investigating committee found the extent of the fraud to be greater than originally believed.

EU defines animals as 'sentient beings'

[LONDON] Member states of the European Union have agreed to change the legal definition of animals to "sentient beings" — rather than goods or agricultural products — when formulating agriculture, transport and research policies.

The change is expected to increase pressure for a ban on the export of live animals. But the new definition is unlikely to affect the way animals are used in research. The union's 15 heads of government agreed the change at last month's intergovernmental meeting in Amsterdam, and a legally binding protocol confirming it will be annexed to the Treaty of Rome, which established the European Community in 1957.

'Protect research time for clinical academics'

[LONDON] The United Kingdom's university vice-chancellors want medical doctors to be given greater incentives to pursue research, amid signs of a slowdown in recruitment to academic posts in departments of clinical medicine. According to the report of a task force set up by the Committee of Vice-Chancellors and Principals, the number of academic staff at the entry-level lecturer grade in the United Kingdom's 21 medical schools fell from 685 in 1991 to 665 in 1996.

The report says pressure from patient care and teaching on the 2,490 serving clinical academics often relegates research to third place in an institution's priorities. Its 35 recommendations include providing protected research time in contracts issued to

clinical academics and improving opportunities for young doctors and dentists to do research during their training.

NIH researcher vows to fight suspension

[WASHINGTON] Edward McSweeney, the researcher at the US National Institutes of Health (NIH) who was recently suspended because of his attacks on the Lyme Disease Foundation, has vowed to appeal against the suspension (see *Nature* 387, 11; 1997). "The public has the right to know that NIH is letting private lobbyists, who are not scientists, interfere with the management of research programs that affect public health," McSweeney said in a statement, adding that NIH had warned him not to talk to reporters.

The NIH suspended McSweeney on 23 June for two weeks without pay, for "insubordination" and "conduct unbecoming a federal employee". A spokeswoman said that McSweeney could file a grievance seeking, for instance, compensatory pay, but that he could not appeal against the decision.

Burying Lenin 'would end unique experiment'



[MOSCOW] A leading member of the Russian Academy of Medical Sciences has claimed that proposals to bury the body of Vladimir Lenin (above), founder of the Soviet state, would end "a unique experiment in preserving human tissues and cells in a constant state for decades".

Yuri Lopuchin claims the international scientific community "would never forgive Russia" for moving Lenin's body from the mausoleum in Red Square, in the heart of Moscow, where it has been kept since Lenin's death in 1924. President Boris Yeltsin has suggested a national referendum on his proposal that Lenin should be buried next to his mother in St Petersburg cemetery, in accordance with his last will.

But Lopuchin argues that Russian techniques of embalming are important for modern science, and should not be lost, because they will be needed "tomorrow if not today" in many fields of human activity. For example, with certain modifications, they could be used for preserving viable genetic material.

Even sociologists need science

Sir — Science is not a cultural construct, not in the strong sense intended by Andrew Pickering. When his book *Constructing Quarks* first came out, I did not take its ‘deconstruction’ of high-energy physics very seriously. However, it is now the object of correspondence in your pages, Pickering is apparently unrepentant, and I feel personally implicated as one of the “opportunists in context” who supposedly infected the high-energy physics community with the quark-gauge theory disease, so I feel impelled to enter the fray.

Like your commentators Kurt Gottfried and Kenneth Wilson (*Nature* **386**, 545–547; 1997), I do not have big complaints about chapters 2 to 13 of *Constructing Quarks*, although I would be more laudatory of the courage of the pioneering Gargamelle neutral-current discovery. These chapters are in general well researched, and I can vouch for details in which I was involved personally, such as the discovery of the gluon. My complaints are rather with the first and last chapters and the arguments Pickering advances in support of his thesis that “there is no obligation upon anyone framing a view of the world to take account of what twentieth-century science has to say”. Since I was interviewed at length by Pickering for his book, I can only wonder whether he delights in being perverse. If he had written a hundred years ago, perhaps his book would have been called *Constructing Atoms*.

This is not the place to contest in detail the conclusions of his book, so I limit myself to taking issue with some of Pickering’s remarks in his recent letter in your columns (*Nature* **387**, 543; 1997). It is simply not true that “the old physics had the merit of explaining almost all of the phenomena that appear in the HEP [high-energy physics] laboratory”.

The old physics was a collection of *ad hoc* recipes to fit limited aspects of the data produced before 1970: it does not explain phenomena seen at the Large Electron–Positron Collider, whereas the new physics explains them almost embarrassingly well.

Pickering vaunts the extensive bibliography of his book; unfortunately, it is out of date and potentially misleading. For example, his footnote 44 refers to five Z-boson events whose “statistical significance... was not overwhelming — recall... the disappearing 6 GeV upsilon”. His potentially deconstructionist readers should perhaps be informed that by now almost 20 million Z particles have been observed, and that their properties agree to great accuracy with the electroweak sector of the Standard Model, which is even better tested than the successful quantum chromodynamics sector advertised by Gottfried and Wilson. By now, the Standard Model has predicted correctly the existence, masses and other properties of the charm quark, the W and Z bosons, and most

recently the top quark: “we got it right” after all.

Pickering is fascinated by the “confusion of war” surrounding the emergence of a revolutionary new paradigm. Indeed, things are not obvious at the start, and one needs intuition and courage to abandon false trails, which provides much of the excitement in doing science. However, nature itself is the only arbiter, and Pickering seems to confuse subjective influences in this initial phase of a scientific revolution with the objective reality of the end result. He suffers from the familiar lack of interest of those studying the sociology of scientific knowledge in the subsequent confirmation of successful new paradigms, and I should also like him to document and compare more thoroughly the rejection by the scientific community of false “discoveries”.

Until he presents a more rounded picture including these elements, the reader is entitled to reject Pickering’s suggestion that “there is no obligation upon anyone framing a view of the world to take account of what twentieth-century science has to say”. My guess is that he himself does take account of it: or perhaps he does not know or care how the Sun shines, does not believe in CD players, and flies in airliners that are navigated by sight?

Jonathan R. Ellis
CERN, 1211 Geneva 23,
Switzerland

More danger in doctrine than in genetics

Sir — In response to recent press coverage of advances in genetics, you are quick to point out the imagined political and social dangers of a belief in biological determinism, citing Nazi Germany as a precedent (*Nature* **387**, 743; 1997).

Perhaps you could explain to your readers why similar comments, quoting the vastly greater number of people who perished at the hands of regimes committed to the dogma of cultural determinism in the Soviet Union, China, Cambodia and elsewhere, are never made. Why is it that the death of millions of Jews is routinely hung around the neck of genetics, while that of millions more belonging to many different ethnic and social groups is never laid at the door of the sociological theories that were used to justify their persecution — particularly when such theories are still widely taught and credited?

Surely, if doctrines were to have health

warnings attached to them objectively assessed by the number of individuals they had harmed, the fashionable Marxist belief that the social environment is much more important than anything else would be rated many times more harmful than any acknowledgement of the influence of genes.

As you say, knowledge is power, but the power to kill millions during our century has come much more from the belief that human beings are the hapless victims of ‘ideology’, ‘society’ or ‘class’ than it has from any knowledge of genetics, however faulty or misapplied it may have been. Your leading article eloquently reveals the extent to which everyone is now acutely aware of the dangers of misapplied genetic knowledge. The real worry is that there is much less awareness of the potentially greater dangers of credulity for doctrines of social-determinism.

Christopher Badcock
London School of Economics,
Houghton Street, Aldwych,
London WC2A 2AE, UK
e-mail: badcock@lse.ac.uk

George Wald believed in Apocalypse now

Sir — John Dowling¹ pays a well-deserved tribute to George Wald’s great contributions to the chemistry of vision.

Wald was also concerned with the disappearance of human beings. He said, in 1975, that he found it difficult “to see how the human race will get itself much past the year 2000”.

Warming to his subject, he could find no reason to believe that his children and his students would “be in physical existence ten, twenty, twenty-five years from now”². In short, “... life on Earth may end as early as 1985”^{2,3}. I am not aware that he retracted this statement.

Thomas H. Jukes
Department of Integrative Biology,
University of California, Berkeley,
Berkeley, California 94720-3140, USA

1. Dowling, J. *Nature* **387**, 356 (1997).

2. Wald, G. *Progressive*, December 1975, pp 22–24.

3. Efron, E. *The Apocalypics* (Simon and Schuster, New York, 1984).

Alphabetical listing and citation rates

Sir — Citation rates are recognized as important in the process of assessing research activity and quality. Questions have recently been raised, however, about the validity of the measure in terms of the unexpected role played by author surname. Previous correspondence¹ produced a series of analyses that appeared to support the hypothesis that there is a relationship between alphabetical position of an author's surname and citation rates. Hence those with a surname initial closer to the beginning of the alphabet can expect higher citation rates than those closer to the end. But there still remains a question about the mechanism by which these results can be explained.

When investigating the hypothesis that surname initial is related to citation rates, our preliminary findings support the previously published work¹. In our analysis of the paper version of the 1994 *Science Citation Index*, citation rates per surname initial were measured by counting the number of columns that each letter occupied. Using non-parametric methods, these values were correlated with a numerical ranking of the alphabet. The resulting correlation showed a moderate

and statistically significant association ($r = -0.425$, $P < 0.05$). This supports the contention that authors with their surname initial at the beginning of the alphabet have higher citation rates. However, these findings cannot be interpreted as evidence of any causal mechanism. In order to assume any causal relation, the condition of isolation must be established, that is, the relationship between the two variables exists when isolated from all other confounding variables. One obvious confounding variable in this case is the naturally occurring differences in the frequency of surname initial². If the letters at the beginning of the alphabet naturally occur more frequently, this is a confounding variable that should cast doubts on the interpretation of the original findings. The effect of confounding variables can be controlled for statistically.

In order to control for the effects of the natural differences in surname initial occurrence, we counted the number of pages occupied by each alphabetical letter in the Greater London Residential Phone Book (1994). Although this is not an exact representation of global surname occurrence, it is an appropriate proxy measure in terms of size and ethnic representation. The correlation between surname initial frequency and citation rates is high ($r = 0.882$, $P < 0.05$). Furthermore,

when surname initial frequency is controlled for in the calculation of the correlation between alphabet position and citation rates, the degree of association is substantially reduced and is not statistically significant. The partial correlation was 0.124 ($P > 0.05$). This indicated that, when isolated, surname initial has no effect on citation rates.

Mark Shevlin

Mark N. O. Davies

Quantitative Research Group,
Department of Social Sciences,
Nottingham Trent University,
Burton Street,
Nottingham NG1 4BU, UK
e-mail: soc3shevlin@ntu.ac.uk

1. Tregenza, T. *Nature* **385**, 480 (1997).

2. Jaekel, K. M. *Nature* **386**, 643 (1997).

CJD transmission

Sir — Your News item "CJD link prompts ban on brain tissue use" (*Nature* **387**, 5; 1997) states that "apparent problems in the communication of scientific information by WHO seem to have been reflected in the fact that an initial version of the press release on the expert meeting on bovine spongiform encephalopathy, CJD and other medicinal products stated mistakenly that 'CJD has been found to be transmissible by

blood'. The statement was later corrected."

The only WHO press release issued after the 24–26 March 1997 meeting on human and animal spongiform encephalopathies was issued on 27 March 1997, and clearly states that "there has been no proven or even probable instance of CJD transmission from human to human by blood transfusion or blood products".

F.-X. Meslin

*Division of Emerging and other Communicable Diseases Surveillance and Control,
World Health Organization,
CH-1211 Geneva 27,
Switzerland*

Greek is the word

Sir—I was shocked to read in Leslie Crombie's book review (*Nature* **387**, 251; 1997) that Greek is dead. I will try to come to terms with the loss of my mother tongue. Nevertheless, I should like to know whether the death was due to apoptosis or to necrosis. Now that I think of it, there is really no answer. These words are now dead.

Andreas Arvanitogiannis

*Concordia University, CSBN, H-1013,
1455 de Maisonneuve Blvd West,
Montreal, Quebec,
Canada H3G 1M6
e-mail: andreas@csbn.concordia.ca*

Don't leave dignity out of the cloning debate

Sir—John Harris clearly doubts that the idea of human dignity is relevant to the ethics of human cloning (*Nature* **387**, 754; 1997). He questions how the intentional creation of a cloned embryo might contravene the notion of human dignity, in a society that accepts both abortion and research on early human embryos. But, although he rightly illustrates the potential hypocrisy of accepting abortion and embryo research at the same time as opposing cloning on the grounds of human dignity, this is hardly a justification for conveniently leaving the question of human dignity out of the debate.

The complex moral issues raised by the prospect of human cloning go to the heart of our self-understanding, our ideas of what it is to be a human being. To deny this would itself be a moral view, and one that would need to be supported by convincing arguments. There really is a fundamental difference between a naturally occurring identical twin, and a child that would be the clone of the person it would look to as its father or mother, and the genetic progeny of the people it would consider to be its grandparents.

Most people would surely agree that an individual should be treated as an 'end' in its own right, and never as simply as a means. I disagree with Harris that this principle "is seldom helpful in a medical or bioscience context", or that "it would outlaw blood transfusions or abortions carried out to protect the life of the mother". People who give blood find the act of donating to be something that enriches their lives, rather than simply reducing them to being the means to someone else's health. And those who value others as ends in their own right are not therefore bound to oppose abortion when essential to preserve a mother's life, since to do so would be treating the mother as a 'means', as much as valuing the child as an end.

Surely it is reasonable to argue that society must debate seriously the implications of human cloning for the individuals who would be created, rather than simply treating the issue as just another potential form of infertility treatment. And in such a debate, the question of human dignity will hardly be inappropriate.

Karim Labib

*Chromosome Replication Laboratory,
ICRF Clare Hall Laboratories,
Blanche Lane, South Mimms,
Potters Bar, Herts EN6 3LD, UK
e-mail: labib@icrf.icnet.uk*

The reform of Russian science

Six years after the end of Communist rule in Russia, attempts to reform Russian science have produced a mixed scorecard. There have been some substantial achievements, but there is still need for further changes.

Boris G. Saltykov

When discussing the relationship between science and society, attention is often focused on the likely effect of the results of research and development on the style and quality of life, at both the national and international level. But of equal interest is the opposite question — how political and economic processes affect the potential of science and the ability of the scientific community to produce new knowledge and promote social progress.

It is from this angle that I will consider some results of a unique social and political experiment — the transformation of Soviet science between 1992 and 1996.

The inheritance

Science in the USSR was an integral part of the economy. The organizational pattern of research and development under Communism was based on the principles and mechanisms of the command-and-control economic system. This system had several characteristics: first, Soviet R&D was strongly militarized. Even as late as the Gorbachev era (the second half of the 1980s), military-oriented R&D spending accounted for about 75 per cent of the federal science budget (in contrast, it has never exceeded 60 per cent in the United States). Similarly, by all other measures, such as the number of researchers or the level of investment in equipment per worker, the resources concentrated in the military R&D sector were well above average.

Second, Soviet science was always kept extremely restricted, so contacts with the international scientific community were limited and strictly supervised. At the same time, many of the results of Soviet research were never made available to the public.

Both of these factors — which can be linked together under a more general principle of requiring both the country's economy and its research to be 'self-supporting' — inevitably led to the emergence of the concept of the need to maintain an 'overall research front'. In other words, the strategy was to carry out research in virtually all existing directions of international science.

The third factor that strongly influenced the effectiveness of Soviet science was its strictly departmental organization. Every research unit was affiliated to a particular ministry or agency — or to the Russian Academy of Sciences — which tightly controlled its research agenda, resources, relations with other economic bodies, and so on. This departmentalization, the dominant



Boris G. Saltykov was appointed Minister of Science and Technology Policy in President Boris Yeltsin's first government in November 1991, and held this position until August 1996. A physicist by training, he is currently director of the Russian House for International Scientific and Technological Cooperation in Moscow.

principle on which Soviet science was organized, led to the unnecessary duplication of research, inhibited flexibility in the allocation of resources, and eventually undermined the way these resources were used.

Under this system, the choice of research priorities was strongly influenced by the priorities (or administrative 'weights') of government ministries and agencies. Furthermore, scientific criteria — and even the very system of values in science — were often replaced in deciding on priorities by the values of the command-and-control system of administration. Managerial obedience and tolerance of the view of 'bosses' were valued more highly than freedom of creative work, and an administrative career, rather than one based on scholarly achievements, became the predominant characteristic of Soviet science.

Another feature of the system was the separation of basic and applied research from the sphere of education. Also, scientific resources tended to be concentrated in large research complexes, in some cases employing thousands of researchers and engineers:

Attempts to overcome the traditional bureaucratic attitudes of ministries and academies have frequently failed

there was a complete lack of smaller institutions. Because salaries in the sphere of intellectual work (as, indeed, throughout the economy) were extremely low, Soviet science developed in a labour-intensive and labour-excessive way. Finally, the Soviet R&D system possessed a feature unique for advanced industrialized economies — it was totally state owned. Everything that happened within Soviet science, including all results of intellectual activities, belonged to the state.

But it would be a serious mistake to concentrate only on the negative aspects of the Soviet pattern of managing science. The system had a logic of its own, and was tailored to meet the principles of the command-and-control economy. It permitted enormous intellectual and technical resources to be focused on selected fields of science and technology, and in this way contributed to impressive achievements in basic research and in military-oriented technologies.

Indeed, many still believe that the former Soviet system, which combined a reasonable degree of freedom in the choice of research topics with lack of any responsibility for the way in which public resources were spent, provided almost ideal conditions for carrying out basic research and development work. The intellectual and educational level of researchers was very high, and some excellent basic research was produced by a system which inevitably relied heavily upon the traditions of pre-revolutionary science in Russia.

But the drawbacks of that system are also obvious. It was wasteful of resources and excessively rigid, and so placed a heavy burden on the state budget. Inflexible planning and administration, as well as a lack of any incentive to reform professional activities, inhibited the movement of researchers between institutions. The system was only able to work smoothly by relying on a permanent influx of new researchers.

Indeed, studies have suggested that Soviet science achieved its most impressive results in the 1960s and 1970s, when the rate of recruitment to the scientific community was between 5 and 7 per cent per year. When the rate of recruitment started to fall in the early 1980s, eventually reaching almost zero, the effectiveness of the way that the system was organized fell dramatically, and Soviet science began to lag behind that in the West.

Perestroika paves way to reform

In the mid-1980s, at the beginning of the period known as *perestroika*, Soviet science was, at least when measured by quantitative factors such as the overall number of

researchers or the share of the gross domestic product devoted to research, comparable to that of the United States. After the collapse of the Soviet Union in 1991, and the subsequent formation of newly independent states, Russia inherited approximately half of the former Soviet Union's population, and a much greater share — between 65 and 70 per cent — of its science resources.

Taking into account only the more efficient parts of the research system, especially basic research and that carried out for military purposes, Russia's share of the Soviet scientific inheritance was even larger. When, at the end of 1991, the government headed by president Boris Yeltsin and prime minister Egor Gaidar came to power in Russia, and the decision was made to move to a market economy and create a democratic state, we faced the problem of drawing up an appropriate strategy for the reform of Russian science. As a result, the era of Soviet science, with all its merits and faults, came to an end in 1991.

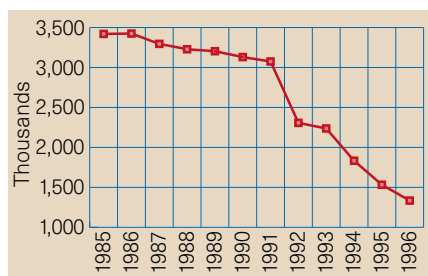
It rapidly became clear that, as soon as the broad demilitarization of Russia's economy began, and the state gave up its responsibility for many types of activity — it stopped paying everyone out of public funds — much science would be left without funds and contracts. Under new economic conditions, and, initially, against the background of a severe economic crisis, the new Russian federal state (which is only one half of the former USSR) would be unable to preserve the whole of Soviet science on its previous scale.

Realistic goals

To cope with the inevitable shrinkage of Russian science, several principles were accepted as the basis of reform. Russia had to give up its commitment to maintaining an 'overall research front' and preserve only those lines of research where we possess world-class abilities. This meant a transition to selective science funding and a careful choice of priorities. It was also essential to secure legal and economic measures to ensure the freedom of scholarly work, as well as publicity for the development of science policy. We also intended to develop the idea that the openness of Russian science was an important principle of its functioning, and to promote its incorporation into the international scientific community.

Finally, one of the most important objectives of the reforms that we introduced was to create a new legal base for science, in particular new regulations concerning the protection of intellectual property rights. All these principles, as well as various other factors, were included in a special decree known as the 'Doctrine of the Development of Russia's Science', enacted by Yeltsin in June 1996.

As far as the creation of modern mechanisms for financing basic research is concerned, the main achievement in recent years has been the establishment of the Russian



Annual average employment in science and science services in Russia, 1985–96.

Foundation for Basic Research, followed by the Russian Foundation for Humanities. At present, 5 per cent of all federal spending on science should by law be distributed through these two foundations, although in practice the figure has been considerably lower.

The administrative patterns of both organizations have been based on that of the US National Science Foundation. Publicly-funded science and technology programmes proved to be an effective lever over scientific and technological priorities. Between 1992 and 1996, financial support was provided to about 40 research programmes in both fundamental and applied fields, ranging from particle physics, biotechnology and research into the human genome to new materials and technologies for processing and storage of foodstuffs.

Overall, the greatest emphasis was placed on the development of the life sciences. Between 1992 and 1994, the number of programmes supported by the Ministry of Science and Technology Policy increased considerably. Unfortunately, however, Russia's economic crisis and other difficulties resulted in a severe underfinancing in 1996, with the result that, by the end of the year, government-funded science and technology programmes had only received 25–30 per cent of the funds that had been promised.

Overall, the shift in science finance has been dramatic. In 1992, the first year of the reform, government spending on science, measured in constant prices, fell by about half, and it continued falling until 1994, when it reached a level of about 30 per cent of what it had been in 1991. In fact, the situation for most research groups was even worse, because the impact of the budgetary cuts was compounded by two other negative factors. One was the drying-up of orders from industry, with the result that government agencies

Russian science has had to give up its commitment to across-the-board strength, and adopt selective funding and clearer priorities

became virtually the only source of finance for research institutions. The second factor was an enormous escalation of the costs of electricity, heat, water and other utilities — which had previously been negligible — that followed the liberalization of prices in 1992.

One result of these various factors was that the average official salary for scientists dropped dramatically to about 65 per cent of the national average in 1992. Later, it grew slowly, and had reached 80 per cent of the national average in 1996. (In the years preceding the reforms, the average salary in science was 10 to 20 per cent higher than the national average.)

Unsurprisingly, this resulted in a significant internal brain drain out of research, with a substantial exodus of scientists and engineers towards new or modernized parts of the domestic economy such as commercial banks, financial and legal companies and the telecommunications industry, as well as general administration and management. Admittedly, this improved the prospects for those who stayed in research. And the brain drain was probably inevitable, since at the beginning of the post-Communist era science and education were the only sources of recruits for the new sectors of Russia's economy. But unfortunately those who left science were often the most energetic and highly motivated young researchers.

The most radical transformations of the Soviet R&D system were those that affected the so-called 'branch sector' of Russian science, where the reduction in the number of both employees and research institutions was substantial. Many institutes, design offices and prototype-engineering organizations were turned into privately-owned enterprises. While the overall number of these institutions decreased by 600, the number of research institutes in both the government and the academic sector grew from 586 in 1992 to 787 in 1995. But much of this was accounted for by the splitting up of large institutions into two, three or even more smaller ones.

A more objective measure is provided by figures for employment in science. By the beginning of 1996, the number of those employed in science and related services had fallen by 57 per cent from its level in 1991 (see graph above). The smallest reduction was in the academic sector — down 17 per cent.

This rapid transformation of the government's research activities made the preservation of Russia's leading scientific and technological institutions a priority. With this in mind, in 1993 we launched a programme to create State Research Centres (SRC). From thousands of R&D organizations, 61 institutes carrying out research in priority fields, such as electronics, aviation, shipbuilding, chemistry and new materials, were selected and given the status of SRC. Between 1993 and 1995, 34–54 per cent of the Ministry of Sci-

ence budget was spent on financing SRC projects. The effectiveness of the SRC programme was demonstrated by an exhibition held in St Petersburg in 1996 of new technologies developed by the centres.

The remaining government-funded laboratories experienced large-scale privatization so that, by 1995, 1,100 R&D organizations — about half of the sector as a whole — had been privatized. Where there have been shortcomings in the privatization of government research, this often appears to have been the result of inadequate methods of assessing the value of intellectual property as a proportion of the total value of privatized scientific and technological organizations.

A new arrival on the Russian research and development scene is a large number of small institutions — many privately run — concerned with promoting innovation. There are now more than 50,000 such institutions, which are becoming real sources of modernization of the national economy. Many of these are supported through a programme set up in 1996 by the State Committee for Science and Technologies to support small innovative business. Even earlier, in 1992–93, the system of state funds for financial support of innovative projects had been established.

Another important development during the post-Communist era has been the laying down of a new legislative framework for science. In particular, a new patent law and a Law of the Right of Authorship have been adopted. In 1996, after three years of heated discussion, the fundamental law 'On Science and State Scientific and Technological Policy' was finally enacted, giving a legal status to most of the new organizational principles on which Russian science is now based, including that of the Russian Academy of Sciences and other academies, and granting intellectual property rights to individuals rather than the state.

The transformation of Russia's scientific community has been greatly influenced by a new policy of openness and support for international cooperation. Not all aspects of this have been positive. For example, Russia has suffered considerable losses because of the brain drain. It has been estimated that between 11,000 and 12,000 scientists and engineers have emigrated from Russia during the past five or six years, and a similar

Under the former Soviet system, managerial obedience and a desire to please the 'bosses' was more important than creative freedom



Helping hand: US Vice-President Al Gore (right) signs an agreement on scientific cooperation with Russian prime minister Victor Chernomyrdin in January 1996.

number are working abroad on (supposedly temporary) contracts.

This exodus is an inevitable, if unfortunate, consequence of Russia's continuing economic crisis. But the effect of the unprecedented growth and increased effectiveness of international scientific contacts should not be underestimated. Russia's new science policy has helped to attract large amounts of foreign aid for our scientists.

The leading role here has been played by George Soros' International Science Foundation, and by aid from European countries through the INTAS programme. Support has also come from many private foundations, including the MacArthur Foundation and the Howard Hughes Medical Institute. Credit must also be given to the support provided to Russia's scientists by foreign learned societies and professional associations.

Another consequence of the openness of present-day Russian science is a significant growth in the number of commercial contracts agreed between Russia's scientific and technological institutions and foreign organizations. The increase in international scientific and technological contacts has made possible a more objective identification of high-quality elements in our scientific activities.

Present-day problems

Many important components of the reforms set out in 1991 and 1992 have now been put into effect, but the transformation of Russian science is not yet complete. The strongest negative factor, which not only hinders the reform of science but also threatens the very existence of science in Russia, is undoubtedly the current budget crisis. The government has virtually excluded science from the list of priorities, as illustrated by the fact that, in 1996 — by far the hardest of the past five years for science — science only received 60 per cent of the funds it had been promised. As a result, many new initiatives, such as state scientific and technological programmes

and state research centres, have yet to become operational.

At the same time, the controlled reform of the network of research institutions has ended in failure, particularly in the academic sector. Ministries and agencies are still keeping many of these alive, even when there are no properly working leaders in institutes, and their equipment is desperately obsolete. At the same time, the use of (limited) funds for the construction of scientific establishments remains far from rational: rather than being concentrated on the most important projects, funds are being distributed across a large number of construction sites, many of which involve projects that span several years. This situation must result in inefficient spending of federal budget resources.

All this stems from the weakness of state-controlled, centralized efforts to implement a genuine science policy and to reform the research system. Attempts to overcome the traditional bureaucratic attitudes of ministries and academies, many of whose officials still cling to a belief in the Soviet methods of management, repeatedly fail.

These are some of the most critical challenges now facing Russian science. But its potential, together with that of the national educational system, remains remarkably high. Acknowledged for the high quality of its basic research, Russia, even in times of deep economic crisis, continues to invest heavily in science, and so in the production of knowledge. Russian universities and institutes still train excellent specialists, many of whom subsequently move to developed countries, including the United States. We are rapidly advancing towards full integration into the world scientific community.

I therefore believe that Russia's scientific community will find both the strength and the ability to overcome its present crisis, to carry through the necessary reforms, and to become, as it was in the past, one of the world's most productive sources of new knowledge and technologies. □

Predicting where we walk

Michael Batty

Models that link human behaviour to the way paths are formed provide excellent predictions of path geometry. In principle, such models can be extended to urban areas, and used to help understand how space affects such phenomena as crime, congestion and pollution.

The application of classical physics to human affairs, known as social physics, grew in popularity through the early twentieth century. It reached a peak in the 1950s and 1960s when gravitation and potential theory was widely used to predict traffic flow, migration patterns and the distribution of population. Since then it has lost its appeal but, suddenly, new ideas from physics based on scaling and self-organization once again appear to have wide applicability to the social sciences.

On page 47 of this issue¹, Helbing, Keltsch and Molnár illustrate a simple model of how walkers, crossing an open space, are able to self-organize their walks into paths or trails which have a geometry highly characteristic of how real paths develop. Their model is based on standard notions of defining a potential for walking at any point within the space, moving walkers towards points of highest potential in the direction they are travelling, updating this potential, and then reiterating the process, adding new walkers randomly, until an equilibrium based on a stable structure of paths emerges.

Walking potential at any point is defined with respect to the current location of any walker. Distance between these locations acts as a deterrent while visibility increases potential. Both are weighted by a measure of attraction of the point that has the potential to be part of a trail. This attraction is referred to as the 'comfort of walking'. It is measured from a baseline environment which is 'uncomfortable' for walking but which is made more comfortable as walkers impress their footprints differentially across the space. The baseline environment, which might be a grassy field, say, gradually grows back through time; thus the comfort of walking is a perpetual fight between the number of walkers who visit each point and flatten the grass into more comfortable paths, and the restoration of the initial environment.

This model, which is based on a potential updated by positive feedback incorporating a logistic limit on the comfort of walking, is quite standard in theories of self-organization². In applications to social systems, particularly cities³⁻⁵, it has frequently been used to model the effects of spontaneous growth such as out-of-town centres or malls,

'edge cities' as they are called⁶, or the rapid rehabilitation of inner cities associated with gentrification, for example.

Helbing, Keltsch and Molnár's model extends these applications in two ways. First, the geometry of the system is an emergent property of the model. In most urban applications of social and statistical physics to date, with the possible exception of fractals⁷, the geometrical structure of the system is assumed — it is an input to the model. Here it is a consequence of the model. One of the longest-standing issues in applying quantitative models to human systems has been the inability of those models to link spatial structure to social behaviour. Here at last is an approach to modelling in which spatial structure emerges as a consequence of behaviour. The implications are profound.

The second significant feature of the new approach is its application to routine spatial behaviour at a fine scale. To date, most applications have been to much more aggregate populations organized into social groups and large spatial units. It has long been assumed that such models, especially those involving movement, would only apply to the prediction of aggregate traffic patterns, but these 'active-walker' models, developed elsewhere in physics, are pitched at the level of micro-behaviour. The emphasis here on pedestrian movements is unusual. By concentrating on

relatively simple behaviour where the goals of each walker are unambiguous (to cross the space from some origin to some destination), and by assuming simple, relatively unobstructed spaces, excellent predictions of the paths taken and the volumes of traffic generated can be made. Figures 3 and 4 on pages 48 and 49 mirror our experience of the way paths develop as a compromise between all distinct walking directions.

These models can easily be extended to deal with more complex situations where terrain is uneven, where directions and visibility across and within the space confound the picture, and where classes of walker with different motivations might mix within the space. However, the real challenge is to extend this type of thinking to built rather than natural environments. Most pedestrian movement of significance in cities is along pavements, and within buildings and public squares. Origins and destinations of walkers are diverse, while the motivation for walking within space has to be extended to more casual pursuits such as resting and strolling, where direction is less important.

But by far the most significant issue involves predicting why certain spaces and routes are less favoured than others. Consider the pedestrian movement patterns shown in Fig. 1 for rooms within the Tate Gallery on London's Millbank. Clearly, some rooms in the gallery are hardly visited at all, and it is difficult to see how environmental conditions could be incorporated into Helbing, Keltsch and Molnár's active-walker model which would generate such a configuration of paths. The underlying conditions do not relate to the space itself but to the configuration of rooms that define the space, to the way they draw people towards them, or repel walkers, and to what lies within each room. These are difficult issues to model but they are essential to an understanding of how space affects a variety of phenomena, such as crime, congestion and pollution, which are highly correlat-

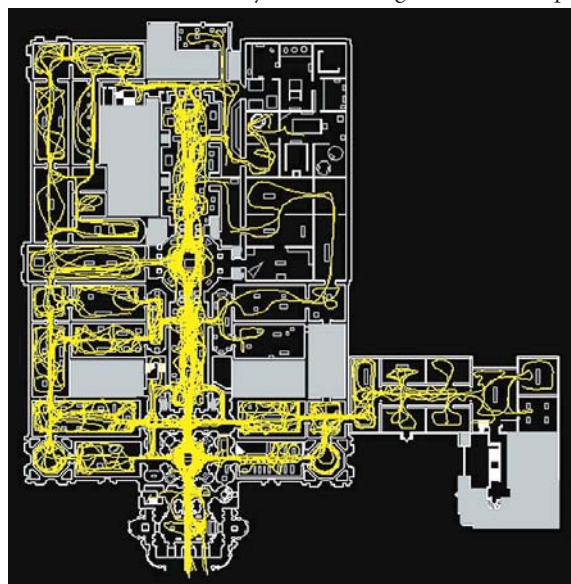


Figure 1 Pedestrian paths based on tracking visitors for the first ten minutes after they enter the Tate Gallery, London (August 1995). The challenge is to develop active-walker models of the kind illustrated by Helbing, Keltsch and Molnár¹, as discussed here, which enable such patterns to be predicted. (Figure courtesy of the Space Syntax Laboratory, University College London.)

ed with movement and intensity of use⁸.

One way of proceeding is to measure the underlying structure of movement which such spaces imply and use these 'movements' to condition the environmental baseline. For example, Hillier and Hanson^{9,10} have developed ways in which the syntax or rules that determine the configuration of such space can be detected through the identification of the most significant lines of sight. These 'axial' lines and their relative connectivity to one another are highly correlated with movement¹¹. They suggest an underlying structure to urban space that is determined by the complexity of buildings or rooms which bound the space.

The syntax based on such lines might be used to condition the environmental variables in the active-walker model, which in turn determine the potential for walking at different points within the space. Further developments for more structured kinds of vehicular traffic at the micro-spatial scale suggest themselves. Traffic prediction along streets is difficult and the macro-models rarely hold up. Active-walker models might be used to embed such macro-predictions into a local geometric context, picking up data on environmental conditions which have no meaning at more macro levels where conventional traffic models operate.

There is a wealth of possible applications and extensions of such models, not only to human but to animal populations, and to social as well as physical spaces. These models naturally extend to problems where the goal is to optimize, to design 'best' paths, as is characteristic of many planning problems where the task is to design high-quality urban environments. □

Michael Batty is at the Centre for Advanced Spatial Analysis, University College London, 1–19 Torrington Place, London WC1E 6BT, UK.
e-mail: m.batty@ucl.ac.uk

1. Helbing, D., Keltsch, J. & Molnár, P. *Nature* **388**, 47–50 (1997).
2. Nicolis, G. & Prigogine, I. *Exploring Complexity: An Introduction* (Freeman, New York, 1989).
3. Wilson, A. G. *Catastrophe Theory and Bifurcation: Applications to Urban and Regional Systems* (Univ. California Press, Berkeley, 1981).
4. Nijkamp, P. & Reggiani, A. *Interaction, Evolution and Chaos in Space* (Springer, Berlin, 1992).
5. Allen, P. M. *Cities and Regions as Self-Organizing Systems: Models of Complexity* (Gordon & Breach, New York, 1997).
6. Garreau, J. L. *Edge Cities: Life on the New Frontier* (Doubleday, New York, 1991).
7. Batty, M. & Longley, P. *Fractal Cities: A Geometry of Form and Function* (Academic, San Diego, 1994).
8. Newman, O. *Defensible Space: People and Design in the Violent City* (Architectural Press, London, 1973).
9. Hillier, B. & Hanson, J. *The Social Logic of Space* (Cambridge Univ. Press, 1984).
10. Hillier, B. *Space is the Machine: A Configurational Theory of Architecture* (Cambridge Univ. Press, 1996).
11. Hillier, B., Penn, A., Hanson, J., Grajewski, T. & Xu, J. *Environment and Planning B* **20**, 29–66 (1993).

Membrane fusion

Bridging the gap by AAA ATPases

Tony Rowe and William E. Balch

Membrane fusion is central to a number of tightly regulated events, including vesicle transport, the maintenance of subcellular compartments, and the disassembly and reassembly of organelles during cell division. Targeting and fusion of transport vesicles¹ requires many membrane-bound and cytosolic factors, and much attention has been paid to the ATPase, *N*-ethylmaleimide-sensitive factor (NSF) and its functional partners — the soluble NSF attachment proteins (SNAPs), and SNAP receptors (SNAREs).

NSF belongs to a growing family of proteins known as the AAA-type ATPases, which are implicated in a diverse array of activities including intracellular membrane dynamics². All of the AAA-type proteins contain one or two copies of a 200-amino-acid ATPase module that is essential for ATP binding and hydrolysis. One member of this family — the product of the yeast cell-cycle gene *CDC48* — is involved in the fusion of endoplasmic reticulum and nuclear membranes³. The rat homologue of *Cdc48p* is p97, and this protein is involved in the reassembly of Golgi cisternae after cell division^{4,5}.

Although much is known about the binding partners for the NSF, the factors that

interact with the other AAA-type ATPases are unknown. But on page 75 of this issue, Kondo *et al.*⁶ report that they have identified an accessory factor, called p47, which is required for p97-dependent membrane fusion *in vitro*. They find that p47 forms a stable cytosolic complex with p97, at a stoichiometry of one trimer of p47 per hexamer of p97. Furthermore, the active cytosolic component that is required for membrane assembly is the p97/p47 complex, rather than p97 alone as was previously suggested^{3–5}.

When examined by electron microscopy, the p97/p47 complex shows a barrel-shaped structure containing a central channel that is ringed by the p47 trimer. This structural organization is very similar to NSF and other members of the AAA-type ATPase family, including molecular chaperones, indicating that they may share similar modes of action. Although the sequence of p47 is unrelated to the SNAPs, it is homologous to yeast Shp1p, which is a protein of unknown function that is required for normal cell growth. So Shp1p is probably the cytosolic binding partner for *Cdc48p*.

With the discovery of p47 by Kondo *et al.*, the molecular details of p97-dependent fusion can now be addressed. For instance,

what is the role of p47? Is it an 'adaptor' protein that mediates the binding of p97 to a membrane-bound receptor(s) — in the same way as SNAPs recruit NSF? And at what step of the membrane targeting/fusion pathway does the p97/p47 complex act? This question is relevant to the debate over the mechanism of NSF action⁷: whereas the original proposal of the SNARE hypothesis was that NSF could promote fusion following membrane docking, recent results indicate that NSF is needed at a pre-docking (or 'priming') step^{7,8}. So NSF, p97 and related AAA-type ATPases may coordinate the protein–protein interactions that mediate membrane priming and targeting⁷, and analysis of the p97/p47-mediated pathway could help to resolve this issue.

Another potentially common feature of membrane-recognition events that involve AAA-type ATPases is the Rab family of GTPases. These proteins are required for many NSF-dependent steps, leading to the homotypic assembly of organelles (that is, fusion of 'like' membranes)^{9,10}, and heterotypic fusion (vesicular transport between different compartments)¹. So are these Rab proteins necessary for the function, in membrane dynamics, of other AAA-type ATPases such as p97? Consistent with this possibility, a Rab GTPase is required for the homotypic assembly of mammalian endo-

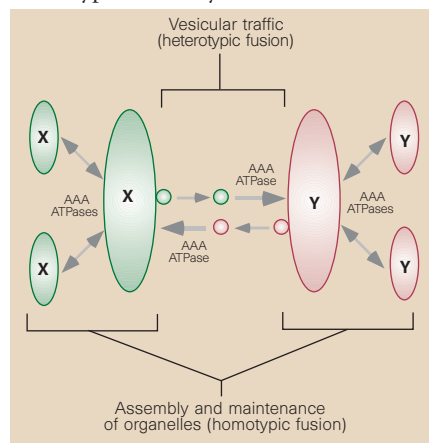


Figure 1 Members of the AAA-type family of ATPases are involved in the membrane dynamics of subcellular compartments. Many of them, such as *N*-ethylmaleimide-sensitive factor (NSF) and yeast *Cdc48p*, have accessory factors. The accessory factor for the rat homologue of *Cdc48p*, p97, has now been found by Kondo *et al.*⁶. p97 forms a complex with this accessory factor (p47), and the p97/p47 complex forms the active cytosolic component for membrane assembly. AAA-type ATPases are required for homotypic events (that is, between 'like' membranes) involved in the assembly and maintenance of different compartments (X and Y), and vesicular trafficking between these compartments. The membrane-recognition events that are controlled by these proteins and their accessory factors may involve an evolutionarily conserved mechanism.

plasmic reticulum cisternae — a process that involves an unknown AAA-type ATPase¹¹.

Even more intriguing is the identity of the membrane-bound p97/p47 receptors. Are they related to the SNARE proteins that are involved in Rab-dependent NSF/SNAP-mediated membrane recognition? If so, AAA-type ATPases may be central to priming the assembly of different membrane-targeting protein complexes.

Why does the cell use distinct, but related, AAA-type ATPases in membrane dynamics? Both NSF and p97 are necessary for reassembly of the Golgi apparatus — NSF seems to mediate the initial heterotypic recognition of small vesicles, whereas p97 promotes the subsequent homotypic-recognition events that lead to the formation of cisternae^{4,5}. Moreover, two AAA-type ATPases (the *PAS1* and *PAS5* gene products²) are required for peroxisome biogenesis in yeast. So the maintenance of subcellular compartments may, in general, involve both homotypic and heterotypic membrane recognition and fusion events, which probably require the activity of diverse AAA-type ATPases (Fig. 1). Given the importance of p97 in the reassembly of the Golgi stack during mitosis, it will be interesting to see whether it, and possibly other, members of the AAA-type ATPase family are regulated in a cell-cycle-dependent manner.

The members of the AAA-type family of ATPases that are involved in membrane

dynamics seem to have ancient origins. Archaeobacteria — cells that lack internal membranes — contain a homologue for p97, suggesting an ancestral role in the fusion events that accompany cell division. Many members of the family could have evolved with the emergence of compartmentalized cells. They may use a subset of structurally and functionally related accessory proteins, and a common molecular mechanism, to promote membrane recognition. But for now, the identification of the p97/p47 complex by Kondo *et al.*⁶ represents an important step towards working out the general principles of how the AAA-type ATPases act. □

Tony Rowe and William E. Balch are in the Department of Cell Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 93037, USA.

1. Rothman, J. E. *Nature* **372**, 55–63 (1994).
2. Confalonieri, F. & Duguet, M. *BioEssays* **17**, 639–650 (1995).
3. Latterich, M., Frohlich, K.-U. & Frohlich, R. S. *Cell* **82**, 885–893 (1995).
4. Rabouille, C., Misteli, T., Watson, R. & Warren, G. *J. Cell Biol.* **129**, 605–618 (1995).
5. Acharya, U. *et al. Cell* **82**, 895–904 (1995).
6. Kondo, H. *et al. Nature* **388**, 75–78 (1997).
7. Morgan, A. & Burgoyne, R. D. *Trends Cell Biol.* **5**, 335–339 (1995).
8. Nichols, B. J., Ungermann, C., Pelham, H. R. B., Wickner, T. & Haas, A. *Nature* **387**, 199–202 (1997).
9. Haas, A., Scheglmann, D., Lazar, T., Gallwitz, D. & Wickner, W. *EMBO J.* **14**, 5258–5270 (1995).
10. Rybin, V. *et al. Nature* **383**, 266–269 (1996).
11. Turner, M. D., Plutner, H. & Balch, W. E. *J. Biol. Chem.* **272**, 13479–13483 (1997).

Chemistry

Oddly ordinary seaborgium

Ron Lougheed

The correct position of new elements in the periodic table is not a trivial matter. We understand the periodic table in terms of the filling of electron shells — a process that would seem to be entirely predictable. But relativistic effects can make the heaviest elements, notably rutherfordium and hahnium¹, behave in unexpected ways. Now, working with just seven atoms, Schädel *et al.* on page 55 of this issue² have extended chemistry to element 106 (seaborgium). They place it firmly in group 6, under tungsten and molybdenum (Fig. 1), raising the question of why it behaves in such a conventional way.

In this region, from the actinides on, one cannot predict properties by simple extrapolation of the periodic table. The high charges on these large atomic nuclei mean that the electrons are tightly bound, orbiting at relativistic velocities (a substantial fraction of the speed of light). And in many cases valence electron levels are close, and their order can be swapped by relatively small effects. Electrons with low angular momentum (in the *s* and *p*_{1/2} shells) overlap the nucleus strongly enough to be drawn into tighter orbits, partly by relativistic effects. That

screens the nucleus electrostatically from other electrons with higher angular momentum, whose orbits consequently expand.

Relativistic effects can even be important in lighter elements — for example, they explain the stability of the +1 oxidation state in thallium — but because these effects increase as the square of the nuclear charge, the greatest effect is expected in the heaviest elements. It isn't easy to predict what will happen, though: the heaviest-known group-5 element hahnium behaves more like the actinide protactinium and the lighter group-5 element niobium than its closest homologue tantalum³; so tantalum could be considered the oddest member of group 5.

If seaborgium is a conventional group-6 element, its properties will be similar to those of tungsten or molybdenum, its lighter homologues; and it could also have some properties similar to those of uranium. Schädel and colleagues have shown by gas chromatography that seaborgium forms volatile oxyhalides similar to the group-6 elements², and by liquid chromatography experiments, using nitric acid and dilute hydrofluoric acid, that it forms neutral

or anionic species similar to the group-6 elements, but not cationic species like uranium. There is no guarantee that other chemical species containing seaborgium won't have more peculiar chemistry — the relativistic effects may be different for different molecular orbitals — but for now the periodic table is firm.

The practicalities of chemistry on these elements are daunting. Chemists have to work with short-lived radionuclides and very low production rates, so experiments must be performed quickly and still efficiently separate the desired nuclide from similar products. Any chemistry that depends on complexes containing two or more of the studied atoms must be avoided, because of the one-atom-at-a-time nature of these experiments — just three daughter decay chains were observed in 5,000 solution-chemistry experiments on seaborgium, for example — and direct measurement of the large number of atoms needed to measure physical properties is impossible. Even though the chemistry of single atoms should be identical to that of a large number, additional experiments with radioactive tracers are usually made to confirm this expectation. Finally, materials or chemicals that act as trapping agents, as many precipitates do, must usually be avoided. (An exception is the clever chemistry⁴ that showed the group-5 element hahnium to be like niobium and protactinium, by its absorption on glass from solution in nitric acid.)

The short-lived man-made elements are produced by charged-particle accelerators, instead of the classical techniques that separate elements from natural materials, nuclear-explosion debris or reactor materials. The particle beam is aimed at a thin target, from which the reaction products emerge at high speed. They are then usually transported in a gas jet to the chemical apparatus in seconds, and separated by gas or liquid chromatography, or solvent extraction.

The subsequent decay, usually by α -particle emission or spontaneous fission, is

1	2	3	4	5	6	7
Rb 37	Sr 38	Y 39	Zr 40	Nb 41	Mo 42	Tc 43
Cs 55	Ba 56	La 57	Hf 72	Ta 73	W 74	Re 75
Fr 87	Ra 88	Ac 89	Rf 104	Ha 105	Sg 106	Ns 107
Actinides			Th 90	Pa 91	U 92	Np 93
Lanthanides			Ce 58	Pr 59	Nd 60	Pm 61

Figure 1 Part of the periodic table, including the early actinides, and the three latest elements to have their chemistry analysed: rutherfordium (104); hahnium (105); and now seaborgium (106).

measured, to determine which chemical fraction contains the atoms (for example, in the liquid chromatography experiments, whether the atoms come through the column or stick to it). These experiments have evolved from a few to several hundred manual operations, with the gradual addition of automation to increase both speed and reliability, until, in the very advanced miniature automated chemistry described in the present paper, even the chromatographic columns are changed automatically, and several thousand experiments are repeated.

To what elements can chemistry be extended? Older predictions were that chemical experiments would end in the heavy actinides. Using present-day production methods and available targets, it is not likely

that chemical studies will extend beyond one or two more elements. Yet new nuclides with half-lives long enough for chemical studies are predicted, at least up to element 114. Only a few atoms of these nuclides might be produced in months of bombardment. Intense radioactive beams using neutron-rich projectiles are one possible route to increased production, still far in the future — but perhaps not so far, when we note that the first chemistry of seaborgium has come 23 years after its discovery. □

Ron Loughheed is in the Lawrence Livermore National Laboratory, University of California, Livermore, California 94550, USA.

1. Schädel, M. *Radiochim. Acta* **70**, 207–223 (1995).
2. Schädel, M. *et al. Nature* **388**, 55–57 (1997).
3. Kratz, J. V. *et al. Radiochim. Acta* **48**, 121–133 (1989).
4. Gregorich, K. E. *et al. Radiochim. Acta* **43**, 223–231 (1988).

Visual neurobiology

Colouring the cortex

Karl R. Gegenfurtner

In humans, the neural basis for colour vision lies in the activity of the ‘colour-opponent’ neurons, which receive inputs of opposite sign from the three different classes of cone photoreceptors that are found in the eye (Fig. 1). Colour-opponent neurons are abundant in the first stages of the visual pathway — the retina and the lateral geniculate nucleus¹. Surprisingly, however, they are observed rather infrequently by

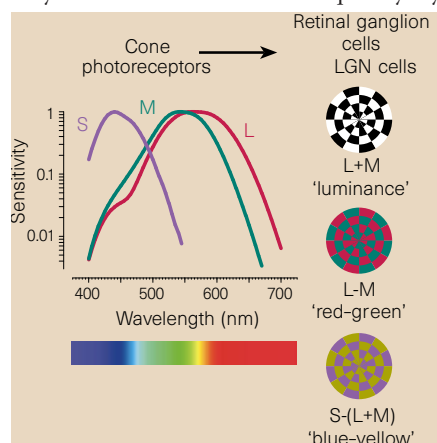


Figure 1 In human colour vision, the information from the three different types of cone photoreceptors, short-, middle- and long-wavelength-sensitive (S, M, L), is combined by retinal ganglion cells and cells in the lateral geniculate nucleus (LGN) into two chromatic channels and a luminance channel¹. The spectral sensitivities of the cone photoreceptors are shown to the left. Chromatic channels receive signals of opposite sign from the cones. These opponent channels are often labelled ‘red–green’ and ‘blue–yellow’, even though the stimuli that differentially activate them (shown to the right) appear to be a noticeably different colour¹¹. Increasing contrast results in the appearance of more saturated colours.

single-cell recordings in the next stage, the primary visual cortex (V1), where neurons that add inputs from all three cone types predominate².

On page 68 of this issue, Engel *et al.*³ report that, in sharp contrast to the single-cell recordings, most of the activity that is detected by functional magnetic resonance imaging (fMRI) of visual cortical areas V1 and V2 in humans is due to the activity of colour-opponent cells. Until now, most fMRI studies have only shown that there is a change in activity in a defined brain region when a particular stimulus is used. But Engel *et al.* have measured differential responses to subtle changes in colour and contrast, so the levels of activity in the fMRI signal can be used to produce response contours for the whole range of colours. By comparing these

responses with behavioural measurements obtained for the same stimuli and observers, they have found that the activity of the red–green and blue–yellow colour-opponent mechanisms seems to underlie both sets of data.

Less than ten years ago, in a survey of neuroscientific methods, Churchland and Sejnowski⁴ exposed a gaping hole in the research — namely, that no method existed to study the activity of large ensembles of neurons, or even whole cortical areas, over a time course of several seconds. The subsequent advances in imaging techniques, and especially the availability of fMRI (ref. 5), promised the missing link (Fig. 2).

Functional MRI is based on blood-oxygen-level-dependent (bold) changes in the magnetic resonance signal. The bold assumption is that the blood oxygen level is an indicator for local cortical activity. Because there is activity everywhere — even in the resting brain — only the activity changes that are induced by stimuli differing in the attribute of interest are measured. The resulting difference in the fMRI signal is typically about five per cent of the total signal, and it depends on the stimulus that is used. For example, if a region responds differentially to a moving stimulus, it is tagged as the ‘motion centre’ in the brain. Several such processing centres have been found, and the trend is going towards a ‘grandmother centre’. This is analogous to the trend towards increasingly specialized processing units⁶, which came about from single-cell recordings during the 1970s. Clearly, it is implicitly assumed that information processing is managed by specialized brain regions — distributed processing goes unnoticed in this standard approach.

More recently, several groups^{7,8} have developed strategies to distinguish between visual cortical areas by exploiting the system-

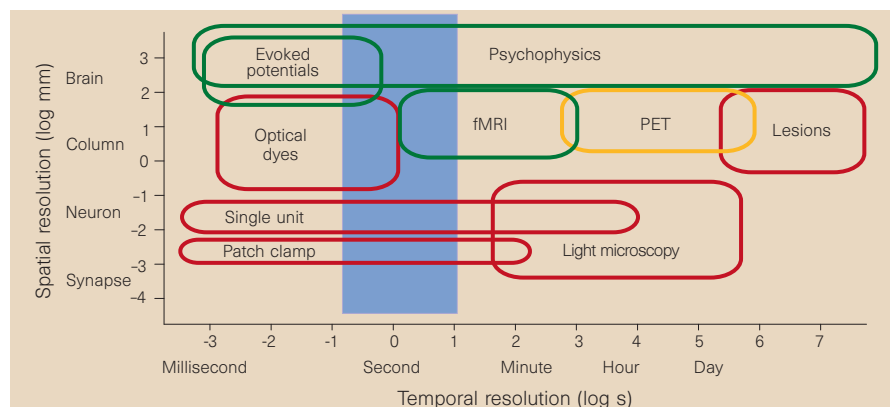


Figure 2 Spatial and temporal resolution of some of the most common experimental methods that are used to study sensory processing. Green boxes indicate non-invasive methods; red boxes indicate invasive methods. The blue shaded area indicates a time slice of major interest, between 100 ms and 10 s. Engel *et al.*³ used functional magnetic resonance imaging (fMRI) to show that most of the activity in the visual cortical areas V1 and V2 in humans is due to the activity of colour-opponent cells. The precise resolution of fMRI is often overestimated — even though images can be acquired every 50 ms with an accuracy of 1 mm, the underlying changes in blood oxygen level are much slower and coarser¹². PET, positron emission tomography. (Modified from ref. 4.)

atic differences in their visual-field representations. In this way, Engel *et al.*³ initially identified V1 and V2, and then determined the colour tuning of these two regions. They used flickering pinwheel stimuli of different contrast (Fig. 1), and measured the fMRI response averaged over the whole of V1 and V2 separately. They found that there was a systematic increase in the fMRI signal with increasing contrast in the stimulus. By repeating this for several different colour combinations, they worked out response contours of equal activity for these brain regions. The result was a map of the activity of two independently identified cortical areas — V1 and V2 — to any colour combination.

What singles out this study is that the fMRI data can be readily correlated with psychophysically obtained measurements. Indeed, the similarity between the psychophysical threshold measurements and the fMRI responses of the same observers to the same stimuli is quite stunning. In both experiments the best responses are obtained for coloured stimuli, the perception of which is mediated by colour-opponent processing. There are also remarkable differences. Most notably, whereas the fMRI data show a bigger response to colour than luminance throughout, psychophysically this holds only for a highly restricted range of stimuli, those with little spatial or temporal variation⁹. Although the fMRI measurements are about 100 times less sensitive than the psychophysical measurements, if we assume that the contours can be directly related to one another, the implication is that there is substantial temporal filtering of the chromatic signal after the V1 region.

The fMRI response to coloured stimuli is also greater than would be expected based on single-cell recordings in Old World monkeys, whose cones and pathways for early visual processing are similar to those in humans. Here, fewer than 15 per cent of all cells give a larger response to colour than to luminance². So why is the fMRI response dominated by the activity of colour-opponent cells instead of luminance cells? To answer this, we need to look at the relationship between single-unit activity and fMRI response. Does the fMRI signal simply reflect the average activity of all neurons? And how specific is the response of an individual neuron to a particular stimulus?

Many of the 'luminance' cells in the V1 region have highly complex receptive-field properties. Their responses are determined by many conjoined properties of the stimulus, including orientation, spatial frequency, temporal frequency, direction and size¹⁰. In single-unit recordings, the idea is to get all of these parameters right for the individual cell, so there will be a strong response to the optimal stimulus but little response to everything else. Because each cell gets its

optimal stimulus, a number of highly active cells are observed. However, it is not clear how many of these cells respond to a particular stimulus, such as the pinwheel used by Engel and colleagues. There might even be an advantage in having a few neurons as possible respond to each luminance stimulus⁶. But the goal could be quite different for colour. Here, the visual system might want to attach a colour to each stimulus, independently of its other characteristics. So it would make sense for colour cells to respond in a less specific manner to these other features.

Although other interpretations of some aspects of these data are possible, the exciting prospect is that we can now study the functional transformation of the response across different visual areas, and correlate the

signals to single-cell responses in these areas as well as to psychophysical responses. □

Karl R. Gegenfurtner is at the Max-Planck-Institut für biologische Kybernetik, Spemannstraße 38, 72076 Tübingen, Germany.

1. Derrington, A. *et al.* *J. Physiol. (Lond.)* **357**, 241–265 (1984).
2. Lennie, P. *et al.* *J. Neurosci.* **10**, 649–669 (1990).
3. Engel, S., Zhang, X. & Wandell, B. *Nature* **388**, 68–71 (1997).
4. Churchland, P. S. & Sejnowski, T. J. *Science* **242**, 741–745 (1988).
5. Kwong, K. K. *et al.* *Proc. Natl Acad. Sci. USA* **89**, 5675–5679 (1992).
6. Barlow, H. in *Cognitive Neuroscience* (ed. Gazzaniga, M. S.) 415–435 (MIT, Cambridge, MA, 1995).
7. Sereno, M. I. *et al.* *Science* **268**, 889–893 (1995).
8. DeYoe, E. *et al.* *Proc. Natl Acad. Sci. USA* **93**, 2382–2386 (1996).
9. Kelly, D. H. *J. Opt. Soc. Am.* **73**, 742–750 (1983).
10. Hubel, D. H. & Wiesel, T. N. *Proc. R. Soc. Lond. B* **198**, 1–59 (1977).
11. Krauskopf, J., Williams, D. R. & Heeley, D. W. *Vision Res.* **22**, 1123–1131 (1982).
12. Maloney, D. & Grinvald, A. *Science* **272**, 551–554 (1996).

Astrophysics

Not-so-cosmic rays

Peter L. Biermann

Cosmic rays are energetic particles that constantly hit the Earth. They include protons and heavier nuclei, along with a few electrons and positrons. It is usually assumed that the flux of cosmic rays is roughly the same throughout our Galaxy, but new data published by Anatolya Erlykin and Arnold Wolfendale may mean that what we see contains the signature of a nearby source, probably a young supernova remnant¹.

Cosmic rays have been seen up to an energy of nearly 10^{21} electron volts (eV) — far beyond any dream of the physicists at the particle accelerators at Stanford, Fermilab or CERN. The flux is very low, from about 10 particles $\text{cm}^{-2} \text{s}^{-1}$ at energies near 10^9 eV, to about 1 particle km^{-2} per century near 10^{20} eV.

More than sixty years ago, Walter Baade and Fritz Zwicky pointed out² that the beautiful nebulae left behind by exploding stars — such as the supernova of AD 1054 observed by Chinese astronomers — could be the powerhouses that accelerate cosmic rays. The standard theory, originally conceived in its basic form by Enrico Fermi about fifty years ago^{3,4}, holds that cosmic rays gain their energy in shock waves created by the explosion of stars. Charged energetic particles gyrate around magnetic fields on both sides of a shock, and bounce off irregularities in these magnetic fields or off turbulence in the gas. Some of these particles bounce back and forth across the shock, gaining energy all the time because the two flows are moving towards each other — compression heats, as a bicycle pump demonstrates.

Three things limit the maximum energy that a particle can gain in this process. First, it can run out of time: the shock gets slower as it

engulfs more material, and therefore less efficient at accelerating particles. Second, it can run out of room: a charged particle gyrating in a magnetic field needs some space for its helical motion, and this space increases with energy (and decreases with the charge on the nucleus), so at a certain energy the gyration can become bigger than the shock wave. Third, it can hit something

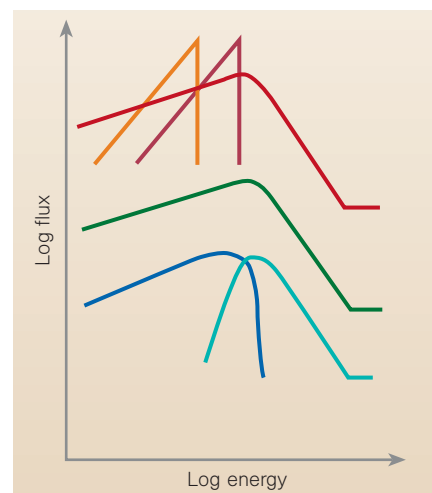


Figure 1 Three knees. The spectrum of cosmic rays hitting the Earth steepens at an energy of about 10^{15} eV (at the 'knee'), and it is relatively easy to explain their production up to about this energy. That could mean: a, that there are two different types of source, which have spectra that join at the knee; or b, that the lower-energy cosmic rays somehow get accelerated, inefficiently, to energies above the knee. But now there is also evidence for bumps, or 'teeth', below the knee, which could be due to different elements accelerated in a nearby supernova remnant. That would mean some changes to either a or b, such as for instance shown in c.

and so lose energy: interactions with ordinary matter and light can brake the acceleration at a maximum energy below the limits imposed by time or space.

Shock acceleration at supernova remnants can explain cosmic-ray energies up to about 10^{15} eV, where, interestingly, the cosmic-ray spectrum suddenly turns downwards (Fig. 1). But we still observe some cosmic rays to energies six orders of magnitude above this 'knee'. Perhaps the higher-energy cosmic rays come from an entirely different source, so the spectrum we see is the sum of two separate spectra (such as the two bottom curves in Fig. 1). Alternatively, acceleration of lower-energy cosmic rays may continue to higher energies, either at shock waves in magnetic stellar winds, or by continuous acceleration in the interstellar medium (by random scatterings from moving magnetic irregularities). In Fig. 1, the central green curve depicts this possibility.

Now the picture may have become more complicated. Erlykin and Wolfendale¹ point out that much of the cosmic-ray flux at Earth could be from the most recent nearby supernova. For cosmic rays coming from a single source, the time and space scales are fixed, so each type of nucleus should have a sharply defined maximum energy, proportional to its charge. This would produce a sawtooth pattern in the spectrum near the knee, each 'tooth' corresponding to an abundant element (Fig. 1, top).

Cosmic rays at these high energies are detected from the 'air showers' that they produce. A cosmic-ray particle enters the Earth's atmosphere and collides with the nucleus of an atom to produce a large number of secondary particles, which hit other nuclei, and so on. At ground level, this translates to a shower of electrons, positrons, photons and muons, whose size depends on the energy of the original cosmic ray. Collating all available data from air-shower experiments, Erlykin and Wolfendale argue that they have discovered 'teeth' corresponding to oxygen and iron near the knee, signs of a recent nearby supernova. Such a supernova would need to be much younger than 10^5 years, and would have to have occurred much closer than 1 kpc. Thus, Vela is a possibility, for instance. If confirmed, this would be powerful evidence for the supernova-remnant theory of cosmic-ray acceleration.

Experimental data now in the pipeline should shed more light on the chemical composition, and thus the origin, of particles with energies from the knee, near 10^{15} eV, up to 10^{18} eV, forming a vital bridge to the highest-energy scales. At the highest energies, above 10^{18} eV, cosmic-ray particles can no longer be contained by our Galaxy's magnetic fields; so as Giuseppe Cocconi realized forty years ago⁵, these particles must come from beyond our Milky Way. It is the goal of the international Auger project — a pair of

detector arrays each covering 3,000 square kilometres, still in the planning stage — to detect these highest-energy cosmic-ray particles. To understand their origin and propagation through the Universe would turn cosmic-ray physics into a tool of fundamental physics at extreme energies. □

Palaeoclimatology

Back at the last interglacial

Joël Guiot

It has been established that, over the past 20,000 years or so, alterations in regional vegetation are controlled by the broad, global pattern of climate change. But what about periods longer than this? Are non-climatic factors such as migration and competition between species involved? The answer to this question is of interest because it can tell us how much we can learn about past climate from the vegetation of the time.

On page 57 of this issue¹, Whitlock and Bartlein describe their analysis of a 125,000-year pollen record of vegetation from Carp Lake in the Pacific northwest of North America. Their conclusion is that climate variations are the cause of the millennial-scale changes in vegetation that they see; that is, the fluctuations between steppe and forest vegetation

Peter L. Biermann is in the Max Planck Institute for Radioastronomy, D-53010 Bonn, Germany.

1. Erlykin, A. & Wolfendale, A. *Astroparticle Phys.* **7**, 1–12 (1997).
2. Baade, W. & Zwicky, F. *Proc. Natl Acad. Sci. USA* **20**, 259–263 (1934).
3. Fermi, E. *Phys. Rev.* **75**, 1169–1174 (1949).
4. Fermi, E. *Astrophys. J.* **119**, 1–6 (1954).
5. Cocconi, G. *Nuovo Cimento* **3**, 1433–1442 (1956).

cover, and between subalpine forest and forest types adapted to warm and dry conditions, are strongly correlated with global climate — see Fig. 2 of the paper, page 58. So low ice volume with maximum summer insolation favoured 'warm-and-dry' forest, and moderate ice volume with minimum summer insolation supported subalpine conifers.

It is then possible to assign vegetation to particular climatic events. So-called stage 5e (the Eemian, the last interglacial before the present one, which ran from about 125,000 to 115,000 years ago) was characterized by warm-and-dry, full forest. Stage 5a (the last temperate period before the ensuing full glacial, 85,000–75,000 years ago) was, by contrast, characterized by a warm-and-dry but less dense forest, composed of both trees

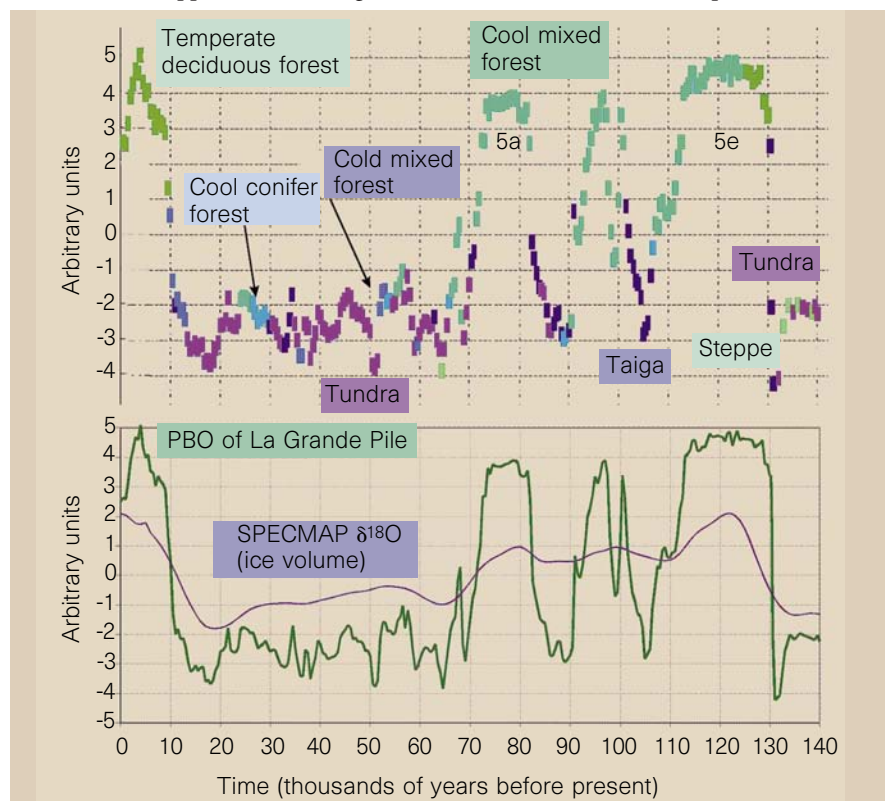


Figure 1 Succession of vegetation biomes in La Grande Pile, a pollen sequence from eastern France¹¹, represented along a timescale (x-axis), and the first principal component of pollen taxa (PBO) describing the main fluctuations (from forest to steppe/ tundra) in the pollen data. The biome succession was determined using the method of Prentice *et al.*¹³. The lower panel also shows ice volume over the same timescale, as represented by the same SPECMAP sequence used by Whitlock and Bartlein¹ (see Fig. 2, page 58).

and shrubs, due to a higher ice volume during a summer insolation maximum. As the authors point out, this last interglacial-glacial cycle provides the possibility of examining the hierarchy of controls of past climate variations.

The notable point about the record analysed by Whitlock and Bartlein is its accurate dating. This was done with a good set of 13 dates from carbon-14 and volcanic-ash records. Moreover, although the pollen-zone boundaries have been assigned independently of the marine oxygen isotope stratigraphy (SPECMAP $\delta^{18}\text{O}$, a proxy for global ice volume), their ages match the isotopic boundaries very nicely.

Accurate dating is, in fact, the crux of the matter, for lack of it has been the main handicap in analysis of the high-frequency fluctuations in vegetation history that are often inferred from continental (terrestrial) cores. Vegetation records are the products of interactions between physical, geological and biological systems, and the difficulty is in relating such records to a given climate mechanism. So it is not surprising that, in recent years, terrestrial palaeoecology has lagged behind palaeoceanography and glaciology, fields in which dating is less problematic and climate forcing can be interpreted more easily.

For example, most of the terrestrial cores contain a lot of high-frequency variations during the last glaciation, but the various explanations for such variations were rarely convincing. Only after glaciologists discovered the so-called Dansgaard-Oeschger events (temperate episodes in the glacial period), and palaeoceanographers discovered Heinrich events² (colder events associated with iceberg discharges into the North Atlantic ocean), has it been possible to give a climatic explanation to the fluctuations in vegetation seen in continental cores. These climatic events have now been recognized in pollen and loess cores from Florida³ to China⁴. The message here is that we need first to understand the climate forcings before being able to interpret terrestrial data, and that these forcings cannot be discovered independently in the terrestrial cores.

Not least, interest in the abrupt climatic events of the past stems from how they might affect us in the future. The present interglacial period (the Holocene, which commenced only about 10,000 years ago) is being analysed to a finer and finer resolution, largely thanks to the wide availability of radiocarbon dates and the existence of annually laminated sediments. Clearly, however, the present interglacial is not yet finished; moreover, records of the course it has taken are sometimes disturbed by man. By contrast, the previous interglacial, stage 5e, gives us the possibility of testing what happens as a glacial period begins by using climate model simulations⁵ and looking at long cores such as that from Carp Lake¹. Such cores are rare, however,

most of them coming from Europe⁶⁻⁹ or South America¹⁰ and often with a timescale dependent on the marine stratigraphy.

The main findings of Whitlock and Bartlein can be generalized to such cores, as illustrated in Fig. 1 by the pollen diagram of La Grande Pile in eastern France¹¹. If we transform this pollen diagram into an index¹² representing fluctuations between forest and herbaceous vegetation, or into a succession of biomes, the correlation with the global climate is evident for timescales higher than 20,000 years — even if, in this case, the timescale has not been derived independently of the ice-volume timescale. The difference between stages 5e and 5a that Whitlock and Bartlein point to can also be seen in this figure. The absence of deciduous forest in stage 5a is likely to be due to cooler winters, which themselves stem from higher ice volume and a cooler Atlantic Ocean.

We need more long sequences, especially in regions which are highly sensitive to climatic changes (for example, the monsoon regions), and analysis of them with sufficient

resolution to detect abrupt events compatible with a human timescale. The climate mechanisms revealed by these records can then be tested with climatic models, so as to help understand the plausibility of future abrupt changes and their effect on the environment. □

Joël Guiot is in the CNRS UA 1152, Faculté de St-Jérôme, F-13397 Marseille cedex 20, France.
e-mail: Joel.Guiot@lbp.u-3mrs.fr

1. Whitlock, C. & Bartlein, P. J. *Nature* **388**, 57–61 (1997).
2. Bond, G. C. *et al.* *Nature* **360**, 245–249 (1992).
3. Grimm, E. C., Jacobson, G. L., Watts, W. A., Hansen, B. C. S. & Maasch, K. A. *Science* **261**, 198–200 (1993).
4. Guo, Z. *et al.* *Clim. Dynam.* **12**, 701–709 (1996).
5. Denoblet, N. *et al.* *Geophys. Res. Lett.* **23**, 3191–3194 (1996).
6. Pons, A., Guiot, J., de Beaulieu, J. L. & Reille, M. *Quat. Sci. Rev.* **11**, 439 (1992).
7. Tzedakis, P. C. *Nature* **364**, 437–440 (1993).
8. Wijnstra, T. A. & Young, R. in *Start a Glacial* (ed. Kukla, G.) 96–105 (NATO ASI Series, Vol. 13, 1992).
9. Follieri, M., Magri, D. & Sadori, L. *Pollen Spores* **30**, 329–356 (1988).
10. Hooghiemstra, H. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **72**, 11–26 (1989).
11. Woillard, G. M. & Mook, W. G. *Science* **215**, 159–161 (1982).
12. Guiot, J., de Beaulieu, J. L., Pons, A. & Reille, M. *Nature* **338**, 309–313 (1989).
13. Prentice, I. C. *et al.* *Clim. Dynam.* **12**, 185–194 (1996).

Environmental engineering

The case of the vanishing beaches



Now (top photograph, 1975) you see it; now (bottom, 1990) you don't. The 'it' is a beach, southeast of Lanikai on the Hawaiian island of Oahu, which disappeared over this period. This and other instances of beach loss or narrowing on Oahu are recorded by Fletcher, Mullane and Richmond in the *Journal of Coastal Research* (13, 209–215; 1997).

The authors identify coastal armouring structures, built to protect shoreline properties from erosion, as the culprits. The trouble is that armouring concentrates erosional forces on the beach directly in front, which, moreover, also loses the replenishment of sand stores from those locked into shoreline land. In all, the authors estimate that about 24 per cent of

the 115 kilometres of Oahu's beaches have been lost or badly affected in this way over the mean (27-year) interval of their study.

Of the consequences discussed by the authors, one is transparent: the possible effects on the Hawaiian economy of a reduction in the number of tourists, who are largely drawn to the islands by the beaches. The 'visitor industry' exceeds by three times that of all others combined in terms of the revenue it delivers to the State of Hawaii.

There are no easy or cheap answers. And as Fletcher and colleagues point out, in a world in which rising sea levels may become an increasing threat, beach loss caused by structures designed to prevent coastal erosion is likely only to increase.

Tim Lincoln

NATURE

100 YEARS AGO

It is well known that the shipments of rum from Demerara, especially during the past year, have been "faulty," and very great pecuniary loss has resulted to the colony. Through the kindness of a friend and the courtesy of the Excise authorities, we received certain samples direct from a bonded warehouse; we were informed that the spirit had been returned at 42 per cent. over proof, equivalent to 74.6 per cent. alcohol by weight. On microscopical examination of a sediment at the bottom of the samples, using a magnification of 1200 diameters, we found chains of small cocci; after the spirit had been kept for some days the cocci were seen to be surrounded with a gelatinous envelope, and after a further interval of time the cocci were found disseminated throughout the liquid, and were rapidly developing and multiplying the observation of the existence and multiplication of any micro-organism in a spirit of such alcoholic strength appears to be of so much scientific interest, and the problem of its presence of such technical importance, that we send this note as a preliminary communication.

From *Nature* 1 July 1897.

50 YEARS AGO

Mr. P. B. Medawar has been elected to the chair of zoology in the University of Birmingham. Not yet thirty-three, he has had a brilliant career at Oxford ... and his research lies along two very different lines. In one he is developing the mathematical treatment of animal form and the process of growth and ageing; in this he carries forward the pioneer work of Sir D'Arcy Thompson. [In the other he is studying] the differences between individuals. Here come his valuable researches on mammalian skin grafting, so important in their human application; his approach to the problem is both immunological and genetical. A recent development has been his discovery and investigation of the curious phenomenon of an induced spread of pigmentation when pieces of black skin are grafted into white areas of a piebald guinea-pig. Prof. Medawar has a fertile imagination combined with an ability for putting his ideas for research into practice. We look forward with confidence to the development of his Birmingham school of zoology.

From *Nature* 5 July 1947.

Signal transduction

Mad about SMADs

Jeff Wrana and Tony Pawson

The transforming growth factor- β (TGF- β) family of signalling molecules can regulate the proliferation of many cell types. Central to TGF- β -mediated signalling in vertebrates are the so-called SMAD proteins and, on pages 82 and 87 of this issue, Hata *et al.*¹ and Shi *et al.*² provide our first glimpse of how the structure of these signalling molecules might regulate their function. Moreover, the new results give an insight into the different ways in which mutations in the SMAD proteins can disrupt their activity in development and human cancer.

For some time, analysis of signalling by TGF- β concentrated on the activation of serine/threonine kinase receptors at the cell surface. TGF- β binds a type-II receptor kinase, which phosphorylates (and activates) a type-I receptor kinase that then initiates downstream signalling. But the discovery of the MAD-related family of signal transducers has strikingly increased our understanding of intracellular events in the TGF- β pathway. The family now includes the prototypic *Drosophila* gene, *Mothers against dpp* (*Mad*), three genes from *Caenorhabditis elegans* (*sma-2*, *sma-3* and *sma-4*) and seven vertebrate members (*Smad1*–*7*). The Smad2 and Smad4 proteins are tumour suppressors, and they are involved in TGF- β -mediated signalling^{3–7}.

Considerable work on the SMAD proteins has fleshed out a general framework to describe how this unique signalling pathway functions. In the case of TGF- β , the activated TGF- β receptor recognizes Smad2 — and probably Smad3 — and phosphorylates them at carboxy-terminal serine residues^{5,7}. Phosphorylated Smad2 then forms a heteromeric complex with another MAD-related protein, Smad4, and accumulates in the nucleus^{6,7}. In the nucleus, Smad2–Smad4 complexes can regulate transcriptional responses by specifically interacting with DNA-binding proteins, such as FAST1 (ref. 8).

An analogous pathway has been described for bone morphogenetic proteins. This pathway involves the phosphorylation of Smad1 and, probably, Smad5, followed by association with Smad4 and nuclear accumulation^{6,9}. So the receptor-regulated SMADs (that is, Smads1, 2, 3 and 5) appear to specify the biological response to TGF- β or bone morphogenetic proteins, whereas their partner — Smad4 — is common to both pathways. The phosphorylation-dependent activation of SMADs provides a simple and direct means by which cell-surface receptors can regulate gene expression.

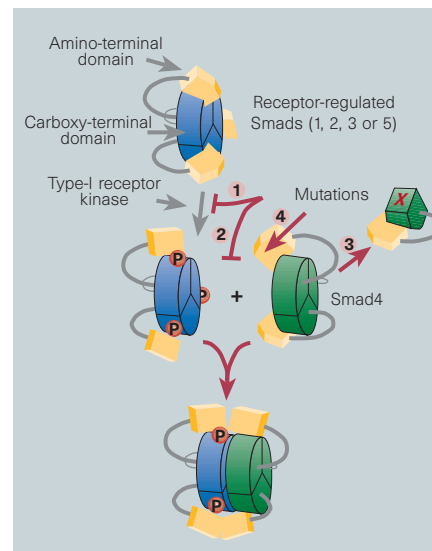


Figure 1 SMAD proteins are involved in TGF- β signalling, and new studies show how the structure of SMADs could regulate their function. The trimeric carboxy-terminal domain of Smads1, 2, 3 or 5 is shown, based on the crystallographic studies of Shi *et al.*². The amino-terminal domain autoinhibits the carboxy-terminal region and Hata *et al.*¹ suggest that phosphorylation of receptor-regulated SMADs (blue) may relieve this inhibition. This will allow formation of a hexameric complex with Smad4 (green), and this complex can interact with DNA-binding proteins to activate transcription. Mutations (red) identified in the carboxy-terminal domain can prevent phosphorylation (1), heteromeric association (2), or trimer formation (3), whereas mutations in the amino-terminal domain may lead to a

It is generally recognized that the amino- and carboxy-terminal regions of SMADs act as regulatory and effector domains, respectively. Hata *et al.*¹ and Shi *et al.*² now suggest potential mechanisms as to how these domains control SMAD function. In their crystallographic studies, Shi *et al.* have solved the structure of the Smad4 carboxy-terminal domain. The monomer is composed of a β -sheet sandwich, capped at one end with a group of three large loops and an α -helix (the loop/helix domain), and at the other by a triple α -helix structure. Interestingly, these monomers assemble to form a trimeric structure in the crystal, with the loop/helix domain of one monomer interacting with the triple helix of the next. So, the overall structure resembles a disk, with the amino termini extending from one face of the disk and the carboxy termini exposed on the side. On the face of the disk opposite to

the amino-terminal side, the third loop from the loop/helix cap is exposed on the surface (Fig. 1). The authors propose that this loop is critical in mediating the formation of a hexameric complex between Smad4 trimers and trimers of phosphorylated, receptor-regulated SMADs.

If TGF- β -mediated signalling occurs through the ligand-dependent assembly of SMAD hetero-oligomers, how is the formation of these complexes regulated? Hata *et al.* provide biochemical evidence that the amino-terminal domain of either Smad2 or Smad4 physically interacts with the carboxy-terminal domain, to prevent the formation of a Smad2–Smad4 complex. Although several models are consistent with these data, the amino-terminal domain could, perhaps, fold back on the carboxy-terminal domain to control its activity, possibly by altering the conformation or sterically hindering the accessibility of the third loop (Fig. 1).

Most of the mutations that affect *Mad*-related genes in cancer cells or in invertebrate development map to the carboxy-terminal domain. Based on structural and functional criteria, Shi *et al.* have now defined three types of substitutions that affect SMAD function. The first class of mutations cause amino-acid substitutions that map to the interface regions between the Smad4 monomers, thereby destabilizing the trimeric complex. The second class map to the third loop on the face of the disk, and these mutations may disrupt the formation of heteromeric complexes. The third class map to the hydrophobic core, and such mutations may destabilize the structure of the protein. But mutant proteins from any of these classes will disrupt the formation of hexameric complexes — so, presumably, they block the transmission of a TGF- β signal.

In contrast to the carboxy-terminal domain, relatively few of the mutations that are associated with a cancerous phenotype have been mapped to the amino-terminal domains of MAD-related proteins. However, two mutations that affect the same amino-terminal arginine residue have been found in both Smad2 and Smad4. Hata *et al.* have characterized these mutations, and they show that they may increase the affinity of the amino-terminal domain for the carboxy terminus. In essence, this locks the regulatory domain onto the effector domain and so prevents signalling of events that would presumably lead to the suppression of cell growth. Such a gain of autoinhibitory activity would represent an unusual mechanism by which to inactivate a tumour suppressor, and it could explain why relatively few missense mutations have been identified in the amino-terminal region of SMADs.

Smad4 does not possess carboxy-terminal serine residues, and it is not regulated by phosphorylation. Yet serine phosphorylation is critical to the function of Smads1, 2, 3

and 5. So what structural role does phosphorylation play in controlling the function of these receptor-regulated SMADs? By analogy with the Smad4 structure, the carboxy-terminal serines of these SMADs are likely to be accessible to the type-I receptor serine/threonine kinase. Phosphorylation at the tail could thus relieve auto-inhibition by the amino-terminal domain, revealing the third loop and allowing the formation of hetero-oligomers (Fig. 1).

Interestingly, many of the mutations characterized within the context of the Smad4 structure were originally identified in receptor-regulated SMADs, and they interfere with their phosphorylation^{4,9}. Because phosphorylation is required for the formation of Smad2–Smad4 complexes, the main defect in these mutants could be in phosphorylation, in the interaction with Smad4, or a combination of both. Solving the structure of full-length SMAD proteins should provide some answers, as well as insights into how the structure of these signalling molecules regulates their function. But remaining

issues include the molecular basis for recognition of specific SMADs by individual receptors, the basis for nuclear accumulation of phosphorylated SMADs, and the precise mechanisms by which SMADs control gene expression. □

Jeff Wrana is in the Program in Developmental Biology and Division of Gastroenterology, The Hospital for Sick Children, Toronto, Ontario, Canada M5G 1X8. Tony Pawson is in the Program in Molecular Biology and Cancer, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada M5G 1X8.

1. Hata, A., Lo, R. S., Wotton, D., Lagna, G. & Massagué, J. *Nature* **388**, 82–87 (1997).
2. Shi, Y., Hata, A., Lo, R. S., Massagué, J. & Pavletich, N. P. *Nature* **388**, 87–93 (1997).
3. Hahn, S. A. *et al. Science* **271**, 350–353 (1996).
4. Eppert, K. *et al. Cell* **86**, 543–552 (1996).
5. Zhang, Y., Feng, X.-H., Wu, R.-Y. & Derynck, R. *Nature* **383**, 168–172 (1996).
6. Lagna, G., Hata, A., Hemmati-Brivanlou, A. & Massagué, J. *Nature* **383**, 832–836 (1996).
7. Macías-Silva, M. *et al. Cell* **87**, 1215–1224 (1996).
8. Chen, X., Rubock, M. J. & Whitman, M. *Nature* **383**, 691–696 (1996).
9. Hoodless, P. A. *et al. Cell* **85**, 489–500 (1996).

Semiconductors

Do the twist to get fit

Pauline Rigby

The substrate problem is a big one for the optoelectronics industry. It is only possible to grow good-quality material, without any defects that would prevent light emission, if a substrate can be found with an almost identical atomic spacing to the layer grown on it. This is lattice matching. For some material systems, such as gallium nitride (GaN), there is no lattice-matched substrate available, so it proved extremely difficult to grow defect-free material, and the development of GaN for blue light emission was delayed for 20 years. To magnify the problem, dozens of different materials are required to cover the wavelength ranges of all optoelectronics applications, and each needs a lattice-matched substrate. But now Ejeckam *et al.*¹ have invented a substrate that doesn't require lattice matching — a universal substrate.

When an epitaxial layer is grown on a substrate with a different lattice parameter, at first the growing layer elastically distorts itself to fit the crystal matrix so that the lattice planes are continuous across the interface. When a certain critical thickness² is exceeded, the elastic energy is released by plastic deformation, in the form of dislocations which nucleate at the surface and then travel in to the interface. A dislocation is a boundary between part of a crystal that has slipped and part that hasn't (Fig. 1). As it moves through the crystal it displaces one row of atoms at a time; on reaching the far side it has displaced an entire layer. The energy barrier to this process is much lower than it would be

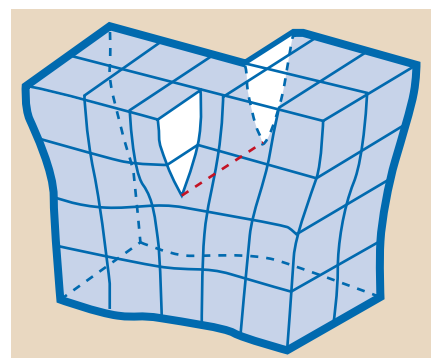


Figure 1 A screw dislocation in a simple cubic lattice.

for all the rows of atoms to move at once.

One approach to the problem is to make the substrate very thin, so extending the critical thickness by sharing the strain between the substrate and overlayer³. As one side of the structure is in tension and the other in compression, the structure bends like a bimetallic strip. These structures are extremely fragile, too fragile to withstand normal wafer handling and processing.

Dislocations are usually detrimental to device performance, but Ejeckam *et al.* have put them to work, to make a stretchy universal substrate. A film of gallium arsenide (GaAs) only 30 Å thick is grown on an ordinary GaAs substrate (with an aluminium arsenide (AlAs) spacer layer in between), then the thin film is bonded to another substrate, essentially by pressing the surfaces

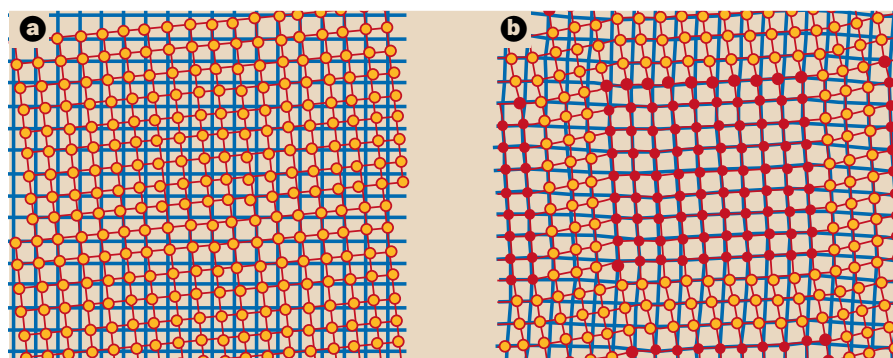


Figure 2 Formation of a twist boundary by rotation about an axis perpendicular to the page. Circles are atoms just above the boundary, and the blue grid marks those below. **a**, The atom positions resulting from a rigid rotation. Very few of the atom positions in upper and lower parts of the crystal are coincident. But a cross-grid of screw dislocations, **b**, permits regions of good match (red) to remain for the same angle of rotation. Around the dislocations the atoms are relatively free to move because the bonds are stretched and therefore weaker. As the angle of twist is increased, more dislocations are needed to accommodate the mismatch, forcing the dislocations closer together and shrinking the regions of good match, so that more of the surface becomes elastic. A layer of another material, with a different lattice constant, can then be grown on top, and the substrate stretches to accommodate it.

together at high temperature. The surfaces must be atomically smooth and chemically cleaned so that the atomic bonds are reformed. After fusing the surfaces together, the original GaAs substrate and the AIs are removed with selective chemical etches.

What makes the universal substrate work is a twist boundary⁴. This is created by rotating the thin film before fusing it to the substrate; in crystallographic terms, introducing an angular misalignment between the $\langle 011 \rangle$ directions of the two lattices. A grid of screw dislocations (Fig. 1) exists in the plane of the interface (Fig. 2) to accommodate the twist. At the cores of the dislocations the atomic bonds are stretched. The final thin film is stretchy rather than rigid because it is sitting on stretched, weakened bonds, and so it puts less stress on any mismatched layer grown on it.

Because the dislocations are fully contained at the interface and don't thread through the thickness of the layer, they don't interfere with the light-emitting properties of the material. To prove the concept, Ejekam *et al.* first grew InGaP on a universal substrate¹. The lattice mismatch was 1%, yet they were able to grow a layer 3,000 Å thick without defects — 30 times the critical thickness.

Then, getting bolder, they tried growing gallium antimonide (GaSb). The lattice constant of GaSb is 6.096 Å, whereas that of GaAs is 5.653 Å — a strain of 8%. Under ordinary conditions defects would start to invade the perfect crystal after four or five monolayers (single-atom layers) had been grown. But the 3,000 Å layer of GaSb grown on a universal substrate is defect-free, despite being more than 300 times the critical thickness (as they reported at a meeting in May^{*}).

A universal substrate made of silicon would be attractive because silicon is the most abundant natural semiconductor, and con-

ducts electricity well. GaAs is the most widely used synthetic semiconductor in the world and it is also an efficient light emitter, unlike silicon, but it has poor conductivity. Using the new technology, it should be possible to combine the best features of both materials on a single optoelectronic chip. Such a chip could use photons to transmit information, while at the same time performing conventional electronic tasks, such as the logic functions that form the basis of computing.

How about a material that is intrinsically capable of both functions? Indium antimonide (InSb) may be such a material⁵: it is luminescent, and has electron and hole mobilities many times better than silicon. The lattice mismatch between InSb and GaAs is 15%, and its critical thickness is less than a monolayer. Nevertheless, a 6,500-Å-thick, dislocation-free layer of InSb has now been grown for the first time, on a universal substrate⁶.

The universal substrate needs to be tested further to determine its limits, but so far it has surpassed the wildest expectations, and has drawn the attention of many leading semiconductor-device manufacturers. The twist principle isn't limited to one type of material, so it could lead to the realization of new classes of compounds, even beyond the realm of semiconductors. □

Pauline Rigby is in the Department of Engineering Science, Oxford University, Parks Road, Oxford OX1 3PJ, UK.

e-mail: pauline.rigby@eng.ox.ac.uk

1. Ejekam, F. E., Lo, Y. H., Subramanian, S., Hou, H. Q. & Hammons, B. E. *Appl. Phys. Lett.* **70**, 1685–1687 (1997).
2. Mathews, J. W. & Blakeslee, A. E. *J. Cryst. Growth* **27**, 118–125 (1974).
3. Lo, Y. H. *Appl. Phys. Lett.* **59**, 2311–2313 (1991).
4. Hull, D. & Bacon, D. J. *Introduction to Dislocations* 3rd edn, 143–144 (Pergamon, Oxford, 1984).
5. Biefeld, R. M. & Hebner, G. A. *Appl. Phys. Lett.* **57**, 1563–1565 (1990).
6. Ejekam, F. E., Lo, Y. H., Seaford, M. L., Hou, H. Q. & Hammons, B. E. *Appl. Phys. Lett.* (in the press).

Daedalus

The watch on the line

Computer database technology is slowly building a subtle tyranny. Already each of us is pinned down by myriad financial and personal records. Security cameras now pose a new threat. Formally, they exist to counter crime. But the growing power of pattern-recognition software will soon allow them to identify and remember the faces and body styles of single individuals. Camera networks will then be able to track every citizen all the time. One possible defence is disguise. False noses, eye patches, stick-on moustaches, joke protruding ears, surgical boots and so on, will become *de rigueur* for all public appearances.

Such dodges, says Daedalus, evade the real issue, which is not surveillance but secrecy. The faceless authorities behind the cameras know all about us; we know nothing about them. Freedom of video information seems a better option. Secrecy would vanish; so would privacy, of course.

In a sense, this is the ultimate logic of the electronic 'global village'. A traditional local village had no privacy; gossip made sure of that. The result was a very low rate of crime and deviancy, and strong social cohesion. The same democracy without privacy might work well for the global village. In this connection, Daedalus recalls his scheme of last week, for coding video signals for easy Internet transmission. So he suggests connecting all security cameras to the Internet.

Crime would plummet. A global audience of eager voyeurs would be far more vigilant than the tired minions who now gaze at the screens. Indeed, many would have a keen personal interest. Motorists would keep an eye on their own cars, parents would track their own children and wives their own husbands, and any criminal caught in the act would probably be recognized by somebody. But profounder consequences could follow. For a start, word of any event worth watching would spread rapidly on the Net. So TV news would be rescued from the selection, editing and slanting of the professionals. And a new form of live entertainment would be born, displacing the endless boring channels of artificial entertainment. Who would bother with a chat show or a soap opera when real crime, sex, disaster, exhibitionism and embarrassment were on offer? And anyone could put on a spontaneous global entertainment anywhere at any time, courtesy of their friendly neighbourhood security camera. Warhol's prediction would come true: we could all be famous for 15 minutes.

David Jones

* Conference on Lasers and Electro-optics, Baltimore, MD, USA, 18–23 May 1997.

Inhibition of ICE slows ALS in mice

Amotrophic lateral sclerosis (ALS) is a progressive age-dependent disease involving degeneration of motor neurons in the brain, brainstem and spinal cord. ALS is universally fatal, with the median survival of patients being five years from diagnosis. In a transgenic mouse model of ALS, we now show that a dominant negative inhibitor of a cell-death gene, the interleukin-1 β -converting enzyme (ICE), significantly slows the symptomatic progression of ALS.

ALS exists in both sporadic and familial forms. Certain familial forms are caused by mutations in the Cu/Zn superoxide dismutase (SOD-1) gene¹. Transgenic mice expressing mutant SOD-1 genes develop an age-dependent progressive motor weakness similar to human ALS². Although little is known about the mechanism of cell death in ALS, downregulation of SOD-1 activity using antisense SOD-1 DNA *in vitro* has been shown to promote apoptosis in a neuronal cell line³. Cell death in this model was mediated in part by the activation of ICE and by binding of endogenously produced mature interleukin-1 β (IL-1 β) to its receptor⁴.

We have previously shown that endogenous mature IL-1 β , produced by activation of ICE, is important in a variety of cell-death models⁵. We recently reported a transgenic mouse expressing a dominant negative inhibitor of ICE, which has the active-site cysteine substituted for a glycine, in neurons under the control of a neuronal specific enolase promoter (NSE-M17Z)⁶. Developmental neuronal cell death does not seem to be inhibited in these transgenic mice, as the brain is of normal size, they exhibit normal

Table 1 Timing of disease onset and mortality

	Mutant		P
	SOD (n=24)	SOD/M17Z (n=22)	
Onset	237.7 $\pm$ 4.4	243.0 $\pm$ 5.6	0.464
Length	11.7 $\pm$ 1.6	27.0 $\pm$ 3.5	0.0002*
Mortality	249.3 $\pm$ 3.8	270.0 $\pm$ 7.0	0.0115*

Disease onset, length and mortality are listed as mean $\pm$ s.e.m. days, illustrating the delay in the mortality of SOD(G93R) transgenic mice by expression of a dominant negative inhibitor of ICE. Asterisks indicate a significant difference (unpaired *t*-test, two-tailed).

behaviour, and have equivalent numbers of neurons in the facial motor nucleus as the wild-type control mice⁶. This contrasts with the NSE-Bcl-2 transgenic mice which have larger brains and more neurons⁷.

To determine whether inhibition of ICE activity *in vivo* might halt the progression of the ALS-like syndrome in mice expressing mutant SOD-1, we crossed five female NSE-M17Z mice from one transgenic founder mouse with one mutant SOD(G93R) male mouse. We determined the genotypes of the progeny from these crosses using the polymerase chain reaction (PCR) to find carriers of the mutant ICE and SOD transgenes, and monitored littermates of SOD(G93R) mice and those with both SOD(G93R) and mutant ICE for the times of disease onset and death. The onset of the disease was scored as the date of the first observation of significantly slower gait and/or limb paralysis. Mortality was scored as the date of death or the inability of the mouse to right itself in 30 s (scorers were unaware of the genotypes of the mice or their birth dates).

Although the timing of disease onset in

the mutant SOD and mutant SOD/ mutant ICE (M17Z) transgenic mice is not different, the double-transgenic mice survive significantly longer from the onset of the disease (27 days) than the mutant SOD mice (11.7 days) (Table 1). Therefore, expression of the dominant negative inhibitor of ICE in neurons of mutant SOD mice is able to slow the symptomatic progression of this disease and delay mortality. Our results indicate that ICE-like proteases might affect disease progression in this ALS mouse model and suggest that ICE inhibitors may be of value in the treatment of ALS in humans.

Robert M. Friedlander

Neurosurgical Service, Department of Surgery,
Massachusetts General Hospital,
Harvard Medical School, Boston,
Massachusetts 02114, USA

Robert H. Brown

Day Neuromuscular Research Laboratory,
Massachusetts General Hospital,
Charlestown, Massachusetts 02129, USA

Valeria Gagliardini, Joy Wang, Junying Yuan

Cardiovascular Research Center,
Department of Medicine,
Massachusetts General Hospital,
Harvard Medical School,
Charlestown, Massachusetts 02129, USA
e-mail: jyuan@bcmp.med.harvard.edu

1. Rosen, D. R. *et al.* *Nature* **362**, 59–62 (1993).
2. Gurney, M. E. *et al.* *Science* **264**, 1772–1775 (1994).
3. Rothstein, J. D., Bristol, L. A., Hosler, B., Brown, R. H. Jr & Kunc, R. W. *Proc. Natl Acad. Sci. USA* **91**, 4155–4159 (1994).
4. Troy, C. M. *et al.* *Proc. Natl Acad. Sci. USA* **93**, 5635–5640 (1996).
5. Friedlander, R. M., Gagliardini, V., Rotello, R. J. & Yuan, J. *J. Exp. Med.* **184**, 717–724 (1996).
6. Friedlander, R. M. *et al.* *J. Exp. Med.* **185**, 933–940 (1997).
7. Martinou, J.-C. *et al.* *Neuron* **13**, 1017–1030 (1994).

Electronic properties of carbon toroids

In a recent Scientific Correspondence, Liu *et al.*¹ reported observing carbon toroids. They suggested that these objects could be considered as quantum wires and that they would exhibit interesting transport properties. Here I provide theoretical support for this proposition.

The ring currents that flow in aromatic molecules in response to the application of a magnetic field are often likened to electrical conductivity^{2,3}, so it is interesting to study the magnetic properties of the carbon toroids. I use as a test case the calculated structure of the C₅₇₆ toroid (ref. 4; and B. I. Dunlap and F. Negri, unpublished). This object is of C_{6v} symmetry and resembles an inflated benzene molecule (Fig. 1). The structure derives most naturally from a [4,4] carbon nanotube⁴. I

used the finite-field London theory⁵ to calculate its magnetic properties. This theory successfully predicted the magnetic susceptibility of C₆₀ (refs 6, 7) and C₆₀[−] (ref. 8).

The toroid has an extremely large and anisotropic ring-current diamagnetic susceptibility, χ_{RC} (Table 1). The calculation clearly shows that the preferred current path is around the six-fold axis. The ring-current magnetic susceptibility of C₅₇₆ with the field perpendicular to the plane ($\chi_{RC}^{\perp}$) is about four times larger than that of the compara-

bly sized, icosahedral C₅₄₀ (which has a ring-current magnetic susceptibility about 13 per cent that of graphite, rotationally averaged, on a per-carbon basis)³. Scaling of $\chi_{RC}^{\perp}$ for toroidal C₅₇₆ with respect to icosahedral C₅₄₀ gives a value of 2.75, based on the number of π -electrons and the mean-square radius of the electronic circulations.

How can we relate the magnetic properties to the electrical conductivity? Following London⁹ we may derive an effective mass (m^*) for the π -electrons of these molecules

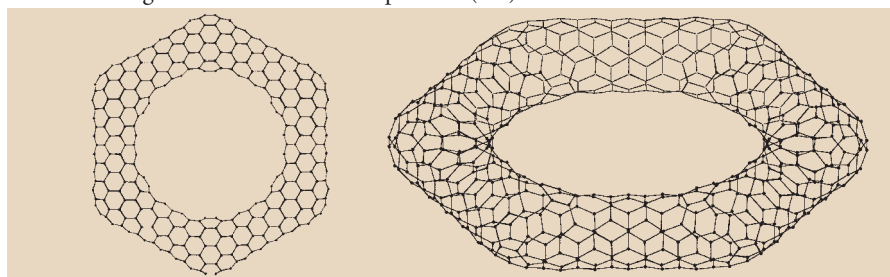


Figure 1 Two views of the C₅₇₆ carbon toroid (after ref. 4).

Table 1 Calculated π -electron ring-current magnetic properties

Fullerene	χ_{RC} (relative to benzene) (p.p.m., centre)	δ_{RC}^*
C ₆₀ (I _h)	-0.5	1.2
C ₅₄₀ (I _h)	144.6	-13.3
C ₅₇₆ (C _{6v} , toroid)	657.2 ($\perp$)	-19.6
	22.4 ()	-1.9

* δ_{RC} is the NMR ring-current chemical shift of a central atom.

by scaling the quantum mechanical result with the free-electron value of the magnetic susceptibility. Using a modern parametrization we obtain an effective mass of $1.5m_e$ for the highest occupied molecular orbital (HOMO) of benzene, whereas a value for the toroid may be obtained from

$$\frac{m^*(C_{576})}{m^*(\text{benzene})} = \frac{\chi_{RC}(\text{benzene})}{\chi_{RC}(C_{576})} \times \frac{\rho^2(C_{576})}{\rho^2(\text{benzene})}$$

$$= \frac{99.8}{4 \times 130.5} = 0.2$$

where ρ^2 is the mean square of the distance of the electronic circulation from the axis of the magnetic field, Z . This is evaluated in the case of the toroid from the atomic coordinates of the carbon atoms:

$$\sum_i^{576} \frac{(x_i^2 + y_i^2)}{576}$$

and thus the area of the toroid in the x - y plane is about 100 times that of the benzene molecule. The ring-current magnetic susceptibility of the HOMO electrons in the toroid is 130 times larger than the total ring-current magnetic susceptibility of the benzene molecule. Thus $m^*(C_{576} \text{ HOMO}) = 0.3m_e$, and because the mobility is usually taken to be inversely proportional to the effective mass we may expect that these carriers will show high conductivities. This is in agreement with the relatively high conductivities measured for linear nanotubes¹⁰⁻¹⁴.

There has been interest in the change in electronic structure that would occur on passing a magnetic flux quantum through a benzene ring. The strong coupling of some of the toroid energy levels to the magnetic field and the concentration of levels near the energy gap (especially in other toroids), suggests that these objects will provide ideal candidates for magnetically induced changes in electronic structure¹⁵. At large size, of course, the level distinctions vanish as the π -electron states merge into energy bands and the flux jumps would be observable as the Aharonov-Bohm effect¹.

R. C. Haddon

Departments of Chemistry and Physics,
University of Kentucky,
Lexington, Kentucky 40506-0055, USA

1. Liu, J. *et al.* *Nature* **385**, 780-781 (1997).
2. Haddon, R. C. *J. Am. Chem. Soc.* **100**, 1722-1728 (1978).
3. Haddon, R. C. *Nature* **378**, 249-255 (1995).
4. Dunlap, B. I. *Phys. Rev. B* **46**, 1933-1936 (1992).
5. Elser, V. & Haddon, R. C. *Nature* **325**, 792-794 (1987).
6. Haddon, R. C. *et al.* *Nature* **350**, 46-47 (1991).
7. Ruoff, R. S. *et al.* *J. Phys. Chem.* **95**, 3457-3459 (1991).

8. Diederich, J., Gangopadhyay, A. K. & Schilling, J. S. *Phys. Rev. B* **54**, 9662-9665 (1996).
9. London, F. J. *Phys. Radium (Paris)* **8**, 397-409 (1937).
10. Langer, L. *et al.* *Phys. Rev. Lett.* **76**, 479-482 (1996).
11. Dai, H., Wong, E. W. & Lieber, C. M. *Science* **272**, 523-525 (1996).
12. Ebbesen, T. W. *et al.* *Nature* **382**, 54-56 (1996).
13. Thess, A. *et al.* *Science* **273**, 483-487 (1996).
14. Fischer, J. E. *et al.* *Phys. Rev. B* **55**, 4921-4924 (1997).
15. Haddon, R. C. in *The Chemical Physics of Fullerenes 10 (and 5) Years Later* (ed. Andreoni, W.) 91-98 (Kluwer, Dordrecht, 1996).

Vision in dim light

On testing my own vision in very dim light I observed two phenomena associated with the lack of retinal rods (the receptors specialized for vision in dim light) in the fovea, the region corresponding to our centre of gaze. First, a bright (or dark) straight line passing through the fovea was seen as discontinuous, with a clear 1° gap. Second, after adapting to dim light conditions, when I blocked light to one eye as far as possible and viewed a brightly lit surface with the other eye, I perceived a swarm of colourless scintillations throughout the visual field of the occluded eye, except for an area about 1° in diameter at the centre of gaze. Each scintillation may represent the simultaneous capture of single quanta by several closely spaced rods.

Damage to a small area of the primary visual cortex produces a circumscribed area of blindness in the visual field, a scotoma, which is not usually apparent as the background fills in the blind area. A straight line crossing through the blind area is generally seen as uninterrupted. This completion phenomenon is also said to occur for lines crossing the normal blind spot produced by the absence of receptors where the optic nerve enters the retina.

The fovea contains cones but no rods. In light that is too dim to activate cones (scotopic conditions) we therefore have a blind spot at the centre of our visual field. I asked whether a line would show a gap when it crossed this foveal scotopic blind spot, or be completed as in the case of a cortical scotoma. On getting up at night I tested this on an abundance of lines, such as the edges of walls, wallpaper designs, and window lattices projected onto walls by street lights. Under these conditions only rods were active (no colours were visible, and any small spots vanished when looked at directly).

I could discern a clear gap when a line passed through the point of fixation. Fixation on a small object is not easy in dim light as the object disappears and gaze tends to wander, but with practice it is possible to fixate on a line so that the gap becomes obvious. Both light lines on a dark background and dark lines on a light background showed a gap as if invaded by the background. Edge boundaries between light and dark areas presented a notch of intermediate brightness. These gaps subtended

about 1° of arc. The corners of a dark or light square were chopped off and similarly replaced by a region of intermediate brightness. In uniformly lit regions, no dark or light spots were seen wandering with shifting gaze. When I fixated on a 12-mm dark spot on patterned wallpaper it disappeared, filled in by the lighter background, and when I viewed a coarsely textured rug I saw a spot of intermediate brightness which wandered to follow my direction of gaze.

I conclude that under conditions of rod vision, line completion does not occur across the foveal scotoma (mentioned in passing in an earlier paper¹), corners are not filled in, nor are patterns and textures completed. But filling-in does occur across diffusely lit surfaces, both dark and light.

When I get up at night, dark adapted over many hours, there is just enough light for me to make my way around but too little to allow me to find small objects. By cupping a hand lightly over one eye before turning on the lights this nuisance can be avoided, because when the lights are turned off again dark adaption is fully preserved in the covered eye, though the other eye is temporarily completely blind. In its thickest part my hand attenuates light by 7-8 log units (measured with a Pritchard photometer, Photo Research PR-1980, Chatsworth).

Recently I noticed a curious phenomenon. With the room light on, facing a bright, uniformly lit wall (luminance $1.8 \log \text{ cd m}^{-2}$), I observed tiny speckles in the dark-adapted covered eye. These bright, colourless points were scattered evenly over the field of view except for an area roughly 1° wide in the centre, and each appeared for a very short time (less than 0.5 s). Their concentration increased dramatically as I let more light into the occluded eye (and tended to be replaced by a wavy, swirling texture), whereas placing two hands over the eye greatly reduced the speckles. They waxed and waned in vividness over a period of many seconds, out of phase with the view seen by the open, light-adapted eye. They were most apparent when I attended to them and faded when attention was switched to the open eye. The variation in brightness is presumably a manifestation of binocular rivalry.

It is well established that the absolute threshold for perception of light is reached when five or more closely spaced rods all capture a quantum of light within a brief period of time². Absolute threshold is commonly estimated at $10^{-6} \text{ cd m}^{-2}$, a level of darkness that may plausibly be reached by my 7-8-log-unit filter. I suggest that the speckles could represent such coincident threshold events, because they appear at very low light levels, they increase in density as more light is admitted, they are absent in the rod-free part of the retina, and they are colourless. I have no idea why they should occur only when one eye is dark adapted

while the other views a bright scene, or what binocular rivalry has to do with the phenomenon.

David H. Hubel

Department of Neurobiology,
Harvard Medical School, 210 Longwood Avenue,
Boston, Massachusetts 02115, USA
e-mail: hubel@warren.med.harvard.edu

1. Hubel, D. H. & Livingstone, M. S. L. *J. Neurosci.* 7, 3378–3415 (1987).
2. Hecht, S., Shlaer, S. & Pirenne, M. H. *J. Gen. Physiol.* 25, 819 (1942).

A polymerase I palm in adenylyl cyclase?

Zhang *et al.*¹ recently reported the long-awaited structure determination of one of the catalytic core domains of eukaryotic adenylyl cyclase, which promises a greater understanding of the regulatory mechanisms associated with the use of cyclic AMP as a second messenger. We have searched the Brookhaven Protein Data Bank² using the program PROTEP³ and found that, far from having a completely novel fold, the fold of the domain bears an extraordinary resemblance to the ‘palm’ domains of the polymerase I family of prokaryotic DNA polymerases, including *Escherichia coli* DNA polymerase I^{4,5} and *Thermus aquaticus* (*Taq*) polymerase^{6,7}. The similarity has important implications for the function and evolution of eukaryotic adenylyl cyclases and related proteins.

The ‘palm’ domain of the polymerases consists of four β -strands and three helices (Fig. 1a). This structure is contained, identical in order and topology, within the adenylyl cyclase catalytic core (ACYc) domain (Fig. 1b). There is no significant sequence similarity between the domains, but 62 α -carbon atoms superpose with a root-mean-square deviation of 1.63 Å (Fig. 1c). An extra domain, the ‘fingers’ domain⁴, lies between strand 1 and helix A of the polymerases (Fig. 1a). In ACYc the amino terminus of the domain and the loop between strand 4 and helix C form an auxiliary lower β -sheet (Fig. 1b) which occupies a similar position in three dimensions to the ‘thumb’ domain of the polymerases.

ACYc does not bind DNA and so the absence of ‘fingers’ and ‘thumb’ is unsurprising. Instead, the lower β -sheet acts as a spacer to create the ‘wreath-like’ dimeric structure of ACYc, in which the ‘palm’-like domains face each other to form a ventral cavity¹. The polymerase I palm domain terminates with helix C, but in ACYc helix C is extended and followed by an additional β -strand and helix (blue in Fig. 1b) and a disordered region. This region has been implicated in regulator binding¹.

The reactions catalysed by ACYc and the

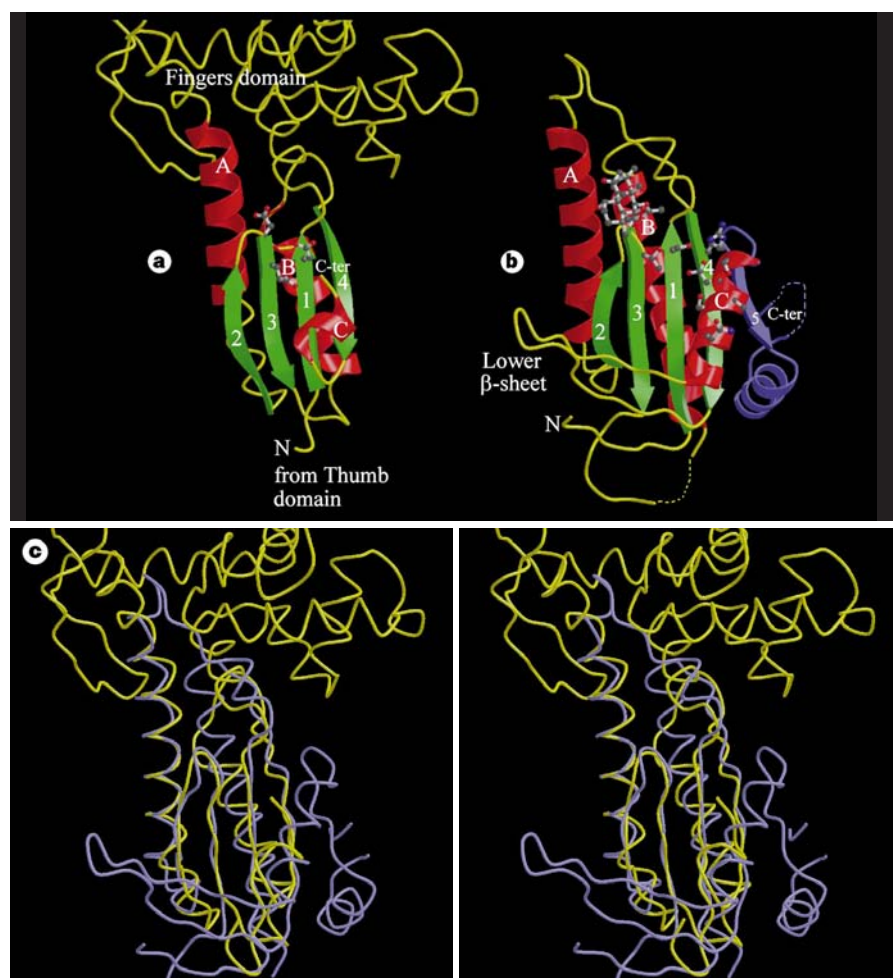


Figure 1 Chain traces^{8,10} of **a**, the palm domain of *Taq* polymerase (ref. 6; PDB code, 1TAQ) and **b**, a monomer of ACYc (ref. 1; PDB code, 1AB8). Equivalent helices and strands, shown as red coiled ribbons and sequentially numbered green arrows, respectively, occur in the same order in both structures. The additional strand and helix in the C terminus of ACYc are in blue. Other non-equivalent parts of the structures are shown as yellow smoothed C α traces, with dashes indicating disordered regions in ACYc. In the polymerase the thumb domain (not shown) is towards the N terminus of the palm domain, and the fingers domain is between strand 1 and helix A. Side chains implicated in the activity of both enzymes are shown in ball-and-stick representation: **a**, left to right, Asp 785, Glu 786 and Asp 610; **b**, top to bottom on helix C and strand 4, Arg 977, Ser 1,032, Asp 1,031, Arg 1,029 and Asn 1,025. For ACYc, a forskolin molecule is shown above Ser 942, Thr 943 and Ser 891 (left to right). **c**, Stereo diagrams⁹ showing the superposition of α -carbons of ACYc catalytic domain (blue) and the palm domain of *Taq* polymerase (yellow) in the same orientation as in **a** and **b**.

DNA polymerases are analogous. Both involve attack by the 3' OH group of a ribose unit on the α -phosphate of a nucleotide 5'-triphosphate, with the elimination of pyrophosphate. However, in the polymerase reaction a deoxyribonucleotide is ligated to a DNA primer, whereas in adenylyl cyclase the reaction involves an intramolecular cyclization within one ATP molecule. The key catalytic residues in the polymerase I active site are three acidic groups (Asp 610, Asp 785 and Glu 786 in *Taq* polymerase) which bind Mg²⁺ and are positioned at the top end of the palm region (Fig. 1a). Although two of the equivalent residues are usually aspartates (Asp 891 and 942) in the C₁-region of ACYc, all three are hydroxyls in the C₂-region reported by Zhang *et al.*¹ (Fig. 1b: Ser 891, which hydrogen-bonds to the highly conserved Arg 977;

Ser 942 which interacts with forskolin at the dimer interface; and Thr 943).

The ATP-binding site in ACYc is not known, but mutagenesis studies^{1,8} show that residues from strand 4 and helix C are involved in ATP binding and catalysis. These residues are adjacent to the positions equivalent to the polymerase active site (Fig. 1a, b) on the same face of the β -sheet, so it seems likely that nucleotide triphosphate will bind in a broadly similar position in both classes of enzyme. Differences in the active site side-chains are to be expected because, in the polymerase, base recognition is in part provided by the complementary template DNA strand. It would be surprising if the detailed chemistry of intra- and intermolecular elimination of pyrophosphate were identical.

These molecules share the same unusual fold, their active sites seem to be in roughly

similar positions and the chemistry that they catalyse is analogous. This suggests an evolutionary link between the families of enzymes, although any common ancestor would be very remote because no sequence similarity remains, and a convergent evolutionary explanation cannot be completely ruled out. Both families are of immense biological importance: DNA and RNA polymerases catalyse reactions that are fundamental to all life-forms, whereas the eukaryotic ACYC core catalytic domains are part of a large membrane-bound protein involved in the more elaborate signal transduction cascades. In view of the extraordinary structural resemblance, we conclude that the adenylyl cyclases may be descended from an ancient nucleotide polymerase ancestor.

Peter J. Artymiuk, Andrew R. Poirrette

David W. Rice, Peter Willett

Krebs Institute,

Department of Molecular Biology and Biotechnology,

and Department of Information Studies,

University of Sheffield, Sheffield S10 2TN, UK

e-mail: p.artymiuk@sheffield.ac.uk

Bryant et al. reply—The adenylyl cyclase fold is a complex three-layer arrangement of α and β structures¹ unlike any in the Brookhaven Protein Data Bank. Two simple and widely distributed motifs, the α/β roll¹¹ and the double split $\beta\alpha\beta$ ($\beta\alpha\beta\beta\alpha\beta$) sandwich^{11,12}, occur as substructures (Fig. 2). The α/β roll comprises part of the inner two layers and the $\beta\alpha\beta\beta\alpha\beta$ sandwich comprises part of the outer two layers. The substructures overlap, each forming part of the central β -sheet. The $\beta\alpha\beta\beta\alpha\beta$ variant found in adenylyl cyclase has strand order 4-1-3-2, by far the most common¹² (the third most common $\alpha+\beta$ structural motif in the Protein Data Bank¹¹), occurring as a stand-alone domain structure in several nucleotide binding proteins and other unrelated proteins¹¹⁻¹³.

The presence of a $\beta\alpha\beta\beta\alpha\beta$ substructure in adenylyl cyclase was identified by us (S. H. B. and J. H. H.) using an automated secondary structure element (SSE) search using the VAST algorithm¹⁴ and 'by eye' (J. J.), and now by Artymiuk *et al.* Structures identified in the VAST search include $\beta\alpha\beta\beta\alpha\beta$ folds in U1a spliceosomal protein (1FHT), procaryoxyphosphatase A (1PCA), ribosomal protein S6 (1RIS), DNA polymerase I (1KFD), and phosphoglycerate dehydrogenase (1PSD), as well as an α/β roll in staphylococcal nuclease (2SOB). These 'hits' are only slightly above the significance threshold of VAST because they involve only 5–7 SSEs. Adenylyl cyclase itself contains 15 SSEs, and when comparing such large proteins, common substructures of this size can often be found by chance¹⁴.

The presence of a $\beta\alpha\beta\beta\alpha\beta$ substructure within a complex fold does not, by itself, support relatedness to any particular $\beta\alpha\beta\beta\alpha\beta$ protein, but DNA polymerase, like adenylyl and guanylyl cyclases, catalyses the

Mg²⁺-dependent in-line attack of a ribose or deoxyribose 3' hydroxyl on the α -phosphate of a nucleoside triphosphate¹⁵. Sufficient similarity in the active-site geometry to test the similarity in functional properties would provide strong support for homology.

The active site of adenylyl cyclase, as defined by mutagenesis, is located in the main ventral cleft between two catalytic domain monomers, possibly adjoining or overlapping one of the regulatory forskolin-binding sites. The forskolin-binding site overlaps with the deoxynucleoside triphosphate (dNTP) binding site of DNA polymerase I¹⁶. In the overlaid structure of the Klenow fragment (1KFD), deoxycytidine 5'-triphosphate (dCTP) approaches two catalytically important residues, Asn 1,025 and Arg 1,029 on the second (non-superimposed) monomer of adenylyl cyclase¹⁷, but has few contacts with the superimposed monomer. None of the polymerase I dNTP-contacting residues are present in adenylyl cyclase. The nucleotides are therefore likely to bind in the same general vicinity, but with quite different intermolecular contacts.

The strongest argument for homology of adenylyl cyclase and DNA polymerase I stems from the possible association of similar structures with a common functional property—the binding of Mg²⁺, which is required for catalysis in both proteins. Ser 891 (β 1) and Ser 942 (β 2– β 3 turn) of the type-II adenylyl cyclase C₂ region overlay with Asp 705 and Asp 882, two of the three conserved Mg²⁺ ligands in DNA polymerase I¹⁶. Homology modelling indicates that there might be an acidic pocket at this position in

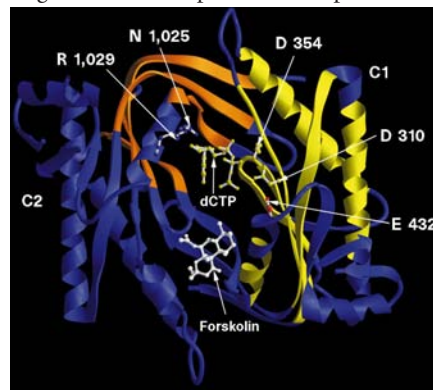


Figure 2 Homology model of C₁ and C₂ catalytic heterodimer of adenylyl cyclase, based on crystal structure of type-II C₂ homodimer. The seven-SSE region overlapping with DNA polymerase I is highlighted in yellow on the C₁ monomer and represents the $\beta\alpha\beta\beta\alpha\beta$ substructure. The seven-SSE region overlapping with staphylococcal nuclease is highlighted in orange on the C₂ monomer and represents the β -barrel portion of the α/β roll substructure. Side-chains in the putative acidic cluster on type-I C₁ are in yellow (Asp 310, Asp 354 of type-I adenylyl cyclase; overlayable with polymerase) and red (Glu 432 of type I; unique to cyclases). Only one of the two forskolin molecules which bind the type-II C₂ homodimer is shown.

the ventral cleft of most adenylyl cyclase catalytic domains, although not in C₂ regions (V. D. Rao and J. H. H., unpublished). If both proteins were to bind the catalytic Mg²⁺ at overlapping sites, the nucleotide substrate would have to bind nearby as well, even if the detailed nature of the sites were entirely different. Structural and mutagenic analysis of Mg²⁺ and nucleotide binding to C₁ and C₂ catalytic domain heterodimers and native cyclases will be essential before definitive comparisons can be made.

Are these similarities less than expected if adenylyl cyclase and DNA polymerase I have evolved from an ancient ancestral gene? Or are they more than expected by chance convergence, given that the $\beta\alpha\beta\beta\alpha\beta$ fold occurs frequently in functionally dissimilar proteins and that imperfect conservation of two potential Mg²⁺ ligands represents limited similarity? Given the biological importance and ancient evolutionary history of both enzymes, the questions are worthy ones, but with our present knowledge we are unable to answer them.

Stephen H. Bryant, Tom Madej

Computational Biology Branch,

National Center for Biotechnology Information,

National Institutes of Health,

Bethesda, Maryland 20892, USA

Joel Janin

Laboratoire d'Enzymologie et

Biochimie Structurales UPR 9063,

Centre National de la Recherche Scientifique,

91198 Gif-sur-Yvette, France

Yu Liu, Arnold E. Ruoho

Department of Pharmacology,

University of Wisconsin School of Medicine,

Madison, Wisconsin 53706, USA

Gongyi Zhang, James H. Hurley

Laboratory of Molecular Biology,

National Institute of Diabetes, Digestive, and

Kidney Diseases,

National Institutes of Health,

Bethesda, Maryland 20892-0580, USA

e-mail: hurley@tove.niddk.nih.gov

- Zhang, G., Liu, Y., Ruoho, A. E. & Hurley, J. H. *Nature* **386**, 247–253 (1997).
- Bernstein, F. C. *et al.* *J. Molec. Biol.* **112**, 535–542 (1977).
- Grindley, H. M., Artymiuk, P. J., Rice, D. W. & Willett, P. *J. Mol. Biol.* **229**, 707–721 (1993).
- Ollis, D. L. *et al.* *Nature* **313**, 762–766 (1985).
- Beese, L. S., Derbyshire, V. & Steitz, T. A. *Science* **260**, 352–355 (1993).
- Kim, Y. *et al.* *Nature* **376**, 612–616 (1995).
- Eom, S. H., Wang, J. & Steitz, T. A. *Nature* **382**, 278–281 (1996).
- Tang, W.-J., Stanzel, M. & Gilman, A. G. *Biochemistry* **34**, 14563–14572 (1995).
- Kraulis, P. J. *J. Appl. Cryst.* **24**, 946–950 (1991).
- Merritt, E. A. & Murphy, M. E. P. *Acta Crystallogr. D* **50**, 869–873 (1994).
- Orengo, C. A. & Thornton, J. M. *Structure* **1**, 105–120 (1993).
- Janin, J. *Nature* **365**, 21 (1993).
- Murzin, A. G., Bressner, S. E., Hubbard, T. & Chothia, C. *J. Mol. Biol.* **247**, 536–540 (1995).
- Gibrat, J.-F., Madej, T. & Bryant, S. H. *Curr. Opin. Struct. Biol.* **6**, 377–385 (1996).
- Koch, K.-W., Eckstein, F. & Stryer, L. *J. Biol. Chem.* **265**, 9659–9663 (1990).
- Joyce, C. M. & Steitz, T. A. *Annu. Rev. Biochem.* **63**, 777–822 (1994).
- Yan, S.-Z., Huang, Z.-H., Shaw, R. S. & Tang, W.-J. *J. Biol. Chem.* **272**, 12342–12349 (1997).

The health of a nation

The Medical World of Early Modern France

by Laurence Brockliss and Colin Jones
Oxford University Press: 1997. Pp. 960. £80, \$150

W. F. Bynum

"They order, said I, this matter better in France", famously remarked the English clergyman and novelist Laurence Sterne as he set out from Dover to Calais on *A Sentimental Journey* (1767). He was referring to his dinner, but we can now judge whether the French ordered their medical care better as well.

The Medical World of Early Modern France is probably the fullest account ever written of a country's medical experience over several centuries. It is certainly the most sophisticated. Laurence Brockliss and Colin Jones have mined an incredible range of archival and printed sources during their 13-year project: manuscript case records and correspondence, matriculation registers and theses, the archives of hospitals and the minute-books of medical corporations, medical advice manuals and learned Latin tomes, diaries and memoirs, the records of court and parliament. The result is a big monograph in every sense: more than 800 pages of text, a bibliography of more than 60 pages and an index just as long.

More importantly, the volume has a clear thematic structure, so that the luxuriant details never lose their significance. Colourful anecdotes and witty asides carry the reader effortlessly along. Despite its magnitude, this is a book that invites reading from cover to cover.

For Brockliss and Jones, the "early modern" extended from the sixteenth century to the eve of the French Revolution. They divided their labours. Brockliss deals primarily with the more formal elements of the landscape — medical knowledge and education, the changing shape of medical theories, the faculties, colleges and institutions that formed the backbone of what the authors call the "corporative" structure of French medicine. Jones covers the more social dimensions of their subject — demography and epidemics, the patient's experience of disease and death, and the medical penumbra of irregular healers, quacks and charlatans, midwives and folk remedies. They have successfully melded their individual writing styles, and the volume speaks with a unified, eloquent voice.

So did the French order their medical system better than their European neighbours? There were certainly differences, for the French developed a highly structured network of corporations during the six-

teenth and seventeenth centuries, which encompassed physicians, surgeons and apothecaries, and was not confined to Paris and the other large cities. It was sanctioned by the state, where the dominant mercantilist economic philosophy gave its blessing to monopolies for those engaged in risky or useful enterprises. The number and power of these medical colleges, faculties and communities declined during the Enlightenment, as those that survived became more élitist but, ironically, less influential.

And French legislation from the late seventeenth century provided a formal framework for medical and surgical training and, by extension, licensing. Surviving records permit Brockliss and Jones to provide reasonable estimates of the number and distribution of physicians and surgeons during the period. By contrast, comparable national legislation in Britain was not passed until 1858.

These formal trappings of "professionalism" (cooperatism might be a more apt word) did not prevent the vigorous activities of drug pedlars and medical showmen, such as Désidério Descombes in the early seventeenth century or Jean Thomas (*le Grand Thomas* was his trade name) a century later. Descombes sold a secret remedy called *orviétan* for a range of routine

maladies such as plague and rabies, while Thomas specialized in pulling teeth. The Pont Neuf in Paris was for a long time a gathering place for quacks and mountebanks plying their wares and services. The policing of such activities was generally spasmodic, and the antiquack literature of French physicians was as lively as that in England, where the medical marketplace was always much more open.

Nor did medical corporatism guarantee peace and harmony among regular practitioners. The medical faculties in Paris and Montpellier long regarded each other with suspicion sometimes bordering on outright hostility; physicians jealously guarded their right to be present at the examination of surgical students; and physicians and surgeons squabbled over which group was best able to deal with the lucrative business of venereal disease.

Despite these and many other examples of intra-occupational conflict, Brockliss and Jones do not conclude that healers in early modern France were simply cynical, self-serving vultures preying on a gullible public. Rather, they argue convincingly that Molière's trenchant critique of doctors in his plays was no more than caricature. For the most part, sick people trusted their medical attendants, and most doctors did



France's famous scientific immortal

This picture of Louis Pasteur seated in his laboratory appeared on an advertising card for the Urodonal Company in honour of the centenary of Pasteur's birth. It is reproduced in the paperback edition of *The Private Science of Louis Pasteur* by Gerald L.

Geison (Princeton University Press, \$16.95, £14.95). Drawing extensively on Pasteur's laboratory notebooks, made available only recently, the book was praised on its original publication by W. F. Bynum in *Nature* 375, 25 (1995).

the best they could. It was not just the doctors who would have seen poetic justice when Molière dropped dead on stage during the third performance of *Le Malade Imaginaire*. (Even hypochondriacs sicken and die sooner or later.)

As always, it seems, medical theories changed more regularly than medical practice. The authors dissect the nuances of Galenism, Paracelsianism, iatromechanism, vitalism and the other 'isms' that permeated medical thinking. They also describe minutely the bleeding and purging that were the mainstays of treatment for many diseases. In 1658 Louis XIV was repeatedly bled and purged during a fever (probably typhus) which almost killed him. Even his physician, Antoine Vallot, was happy, Ambroise Paré-like (*Je le pensay, Dieu le guarit*), to attribute the outcome to providence. (The 20-year-old monarch lived for another 57 years.) Both doctors and patients approached their mutual encounters with a healthy dose of fatalism.

In 1720, five years after Louis' death, plague struck France for the last time. Brockliss and Jones analyse the epidemic in Marseilles and argue that it was a turning-point for the country's doctors. Although the reasons why the plague was confined to Marseilles were more administrative than medical, the doctors received much of the credit. And they earned it: almost 22 per cent of the doctors sent to care for the plague victims in 1720 died. Traditional advice — medical and lay — about how to cope with plague was contained in the aphorism *cito, longe, tarde* (leave early, go far, come back late). By remaining at their posts, doctors had demonstrated their integrity, even courage.

There are many other facets to this wonderful book: vivid reconstructions of the contours of several medical practices; analysis of medical incomes; reflections on

the development during the Enlightenment of preventive medicine, hospitals and surgery; shrewd insights on the changing shape of medical ethics and doctor–patient relationships. These are but a few examples. The volume is quite simply a stunning achievement, a monument to cooperation and humane scholarship. □

W. F. Bynum is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BE, UK.

The higher idiocy

Yes, We Have No Neutrons: An Eye-opening Tour Through the Twists and Turns of Bad Science

by A. K. Dewdney

Wiley: 1997. Pp. 180. £17.99, \$22.95

Walter Gratzer

"Anybody who goes to a psychiatrist ought to have his head examined," Sam Goldwyn sagely observed. Here at least was one central European (born Shmuel Gelbfisch) who was not in the grip of the prevailing fashion. The tide has since turned against Freud (not before time), and for Professor Dewdney in his pursuit of intellectual deviance he presents a soft target.

The subtitle of Dewdney's book rather begs the question of what anyway constitutes science. A definition is always elusive, and the term has in any case been stolen from us, for what greater oxymoron can one conceive of than 'political science'? No doubt we shall one day hear talk of literary science and perhaps even theological science; after all, we already have creation science. So today's jest becomes tomorrow's degree course. But Dewdney, who seems to base his own valuations entirely on the teachings of Karl Popper, might have recalled that the sage regarded psychoanalysis, which he loathed, as the

antithesis of science. It is not bad science: it is not science at all.

Dewdney keeps clear, rather surprisingly, of the 'dismal science' of economics, which John Maynard Keynes, unlike a horde of his successors, regarded as unamenable to quantitative analysis. Nor, in the social and medical line, has Dewdney considered one of the most truly pernicious examples of spurious science — the frontal lobotomy operation, which for decades was inflicted indiscriminately on the vulnerable, most often without consent and with almost uniformly disastrous consequences. For this Egas Moniz, who invented the procedure, was rewarded with the 1949 Nobel Prize for Medicine.

Dewdney does deal effectively with the measurement and heritability of intelligence, and the racial undertones of that sinister subject. But his adherence to Popper, with frequent incantations about what is and what is not falsifiable, leads him to an extreme view of, for example, the search for extraterrestrial intelligence (SETI). In truth this is no more "bad science", as Dewdney defines it, than the study of cosmology or of evolution, where one has to make do without controls, and the principle of falsifiability scarcely helps. Dewdney, in common with many others, is entitled to regard SETI as a waste of public funds and certainly one of the more chancy ventures. But, Popper notwithstanding, an intelligent signal from the void would be sufficient to prove him wrong.

Dewdney lumps SETI with various failed initiatives, driven by manic ambition, greed or injured pride, such as the cold-fusion fiasco and the ill-starred Biosphere 2 project. But can one really compare neural networks with these sad failures? Dewdney does not assert that the many successes of this computational technique are all illusory, only, when one reads carefully, that it has been hyped far beyond its real value. That may be so, but by such a measure how much of recent cancer research could escape damnation (let alone gene therapy, say)? I would have chosen catastrophe theory as a better example in applied mathematics; this enjoyed a short-lived vogue a decade or so ago, with supposed solutions to half of all human predicaments. Among many other revelations that stick in the mind, it presented an astonished world with a rigorous proof that an aggressive dog would attack suddenly, rather than gradually.

Dewdney, who is a mathematician and who inherited the "Mathematical Recreations" column in *Scientific American* from Martin Gardner, is best on his home ground. He is especially clear on issues of probability (SETI) and statistical analysis (IQ tests). He writes in a wearily populist style and is profligate with clichés; nor is he too fastidi-

New in paperback

Radiogenic Isotope Geology

by Alan P. Dickin

Cambridge University Press, £29.95, \$44.95

"Fills an important niche by providing an encyclopaedic, research-level review of the principles, methods and applications of radiogenic isotopes in geology... a notable contribution to the scientific literature on radiogenic isotope geology". Joel D. Blum, *Nature* 381, 126 (1996).

An Interpretive Introduction to Quantum Field Theory

by Paul Teller

Princeton University Press, \$16.95, £14.95

"This is a valuable book. Indeed, the specific interpretive proposals [Teller] makes could have an influence on hidden-variable theories....

[The] book reflects a welcome trend toward a reassessment of the notions of meaning and visualization in contemporary science". Peter Holland, *Nature* 378, 454 (1995).

Resistance to New Technology: Nuclear Power, Information Technology and Biotechnology

edited by Martin Bauer

Cambridge University Press, £24.95, \$39.95

"Consists of papers presented at a conference at the Science Museum in London. Under the rubric of 'resistance', many different stories are told from many different perspectives. Some are stories of comparative diffusion; others of policies for regulating particular technologies; some of public attitudes to certain new technologies". David Edgerton, *Nature* 376, 653 (1995).

ous when it comes to syntax. "Spearman, who treaded where angels evidently feared to..." is how one of his less inviting sentences begins.

Dewdney might also have taken more trouble to check facts. He misspells consistently the name of one of the two principal protagonists in the cold-fusion story (and quite a number of other names besides), bestows a knighthood on Charles Darwin and credits Lynn Margulis, not James Lovelock, with inventing the Gaia hypothesis. He would have us believe that tritium is "a tri-atomic molecule" and that the emission spectrum of hydrogen consists of "dark and light bands". He has Paul Broca joining in the pursuit of the chimerical N-rays 20 years after his death.... I could go on.

But perhaps none of this matters much if the message — that unreason stultifies, costs money and sometimes even kills — gets through to the general reader, for whom presumably the book is intended. The trouble is that, as the writer Franz Werfel put it, for those who believe, no explanation is necessary, while for those who do not believe no explanation is possible.

Walter Gratzer is at the Randall Institute, King's College, University of London, 26–29 Drury Lane, London WC2B 5RL, UK.

Riddles in the sky

Unsolved Problems in Astrophysics

edited by John N. Bahcall and Jeremiah P. Ostriker

Princeton University Press: 1997. Pp. 377.
\$69.50, £55 (hbk); \$24.95, £19.95 (pbk)

Wendy Freedman

The Vela satellites made a serendipitous discovery while watching for illegal nuclear explosions in space in the 1960s. Gamma-ray bursts were detected whose origin has remained one of the unanswered questions in modern astrophysics. These bursts involve the release of about 10^{50} ergs of radiation into small volumes of the order of 1,000 km across with short durations of hundreds of seconds down to a few microseconds.

Hundreds of gamma-ray sources continue to be discovered each year, but even the basic issue of whether they are of Galactic or extragalactic origin has remained unresolved for more than a quarter of a century.

These sources are among the enigmas discussed in this compendium of unsolved problems edited by John Bahcall and Jeremiah Ostriker. And now, within months of the publication of this volume, evidence appears to have been found confirming that these gamma-ray bursts are of extragalactic origin (see *Nature* 387, 859; 1997).

Nevertheless, as this lively volume attests, there are still many exciting prob-



The lure of the red planet

The "grand canyon" on Mars (above) stretches 3,100 miles and extends to a width of 62 miles, dwarfing Earth's Grand Canyon which is a mere 217 miles long and 11 miles wide. From

Destination Mars: In Art, Myth, and Science by Martin Caidin and Jay Barbre with Susan Wright (Penguin, \$29.95), an illustrated survey of the enduring fascination of the red planet.

lems left to interest the graduate student entering the field of astrophysics. The articles are based on invited talks given at a conference held at the Institute for Advanced Study in Princeton in April 1995 in honour of Bahcall's sixtieth birthday.

But the editors have made a great effort — successfully in my opinion — to address a very different audience. The written version is aimed at graduate students at the beginning of their courses, and should also be of interest to senior undergraduates. Each essay was apparently written with this question in mind: "I am thinking of doing a thesis in your area. Are there any good problems for me to work on?"

The present time is perhaps unparalleled in astrophysics. A healthy confrontation between theory and observation exists in many diverse areas and, as new instruments, telescopes and surveys reach fruition, the prospects over the next 10 to 15 years of gaining a deeper physical understanding of the Universe are enormous.

For example, mapping of the anisotropies in the cosmic microwave background down to arcminute scales will give a glimpse of the earliest structures in the Universe and lead to a measure of the geometry and other key parameters of cosmology. Diverse methods are being applied to measure the expansion rate and the mean mass density of the Universe, and to establish lim-

its on the amount of energy density in the vacuum of space. Large surveys of galaxies, active galactic nuclei and quasars, both local and distant, will provide data to confront increasingly detailed theories about the formation and evolution of galaxies and clusters of galaxies.

Our current view is that the dark matter that makes up most of the mass in the Universe is not composed of ordinary baryons. Continuing efforts to elucidate the nature of this matter range from laboratory cryogenic detector experiments searching for weakly interacting particles to observations on the largest scales aimed at measuring the mean mass density and its distribution.

Finally, coming decades are likely to shed light on the "dark ages", the redshift interval between 5 and 1,000, corresponding to the time interval 10^6 – 10^9 years after the Big Bang, where no observations have so far been available.

Each article contains excellent bibliographies and notes suggesting entry points to the literature. I would have loved to have read such a volume when I was a student starting off in the field. It also provides an excellent summary of exciting areas in astrophysics of interest to a wider readership than just the students for whom it is intended. I would recommend it to interested physicists at all levels, and also as a good conference summary proceedings with review talks on a

variety of subjects.

Although the volume gives only a selection and does not cover all the areas of current interest in astrophysics — there is no chapter on quasar absorption lines, for example — students will find it an excellent resource to stimulate their discovery of a rich and exciting field.

The influence of Bahcall can be felt throughout the book, and not only in an editorial sense. Beyond his own direct contributions to many fields in astrophysics (including solar neutrinos on which he provides an article), many authors here attest to the impact he has had in encouraging those around him to choose fundamental problems, to bear in mind the observational consequences of theoretical speculations, to push theoretical consequences to their limits, and to apply advanced analytical and statistical tools. Many visitors to the Institute for Advanced Study have encountered the high standards of excellence set by Bahcall during Friday “stuffy lunch” presentations. □

Wendy Freedman is at the Carnegie Observatories, 813 Santa Barbara Street, Pasadena, California 91101, USA.

Science and salvation

Redeeming Culture: American Religion in an Age of Science

by James Gilbert

University of Chicago Press: 1997. Pp. 390. \$28.95

George Marsden

Early in the twentieth century, sociologists predicted that, with the growth of science and technology, traditional religion would decline. In the United States that has not happened (see *Nature* 386, 435; 1997). Rather, science and technology flourish alongside every variety of religion, from traditional to New Age. The group suicide of the Heaven's Gate on-line space-age cult is just the latest reminder that almost any sort of scientific-religious combination can emerge in that fertile environment. Even among academics, natural scientists in America are more likely to be traditionally religious than are humanists.

James Gilbert's *Redeeming Culture* provides some fascinating background for understanding the interactions of science and religion in the United States. Starting with the Scopes trial of 1925, which shaped an image of antagonism between science and biblicist Christianity, Gilbert describes about a dozen major encounters between American science and religion, mostly from the era just after the Second World War. His account is admittedly episodic, rather than

systematic, but it does provide intriguing pictures of some of the highlights in this cultural exchange, not only among intellectuals, but also at the popular level.

One especially revealing incident is how in the early 1950s the US Air Force regularly offered its troops Moody Bible Institute science films such as *God of Creation* and *God of the Atom* as part of their moral training. Another is the immense mainstream publicity and then controversy surrounding Immanuel Velikovsky's bestselling *Worlds in Collision* which alleged that a crash of a comet into Mars helped to validate events in the Hebrew scriptures such as the Sun standing still. Gilbert also recounts popular efforts to relate science and religion, as in the films of Frank Capra or at the Seattle World's Fair of 1962.

Although Gilbert does not argue any strong thesis about the overall significance of these episodes, they do suggest that the cultural status of discussions of science and religion were very different half a century ago than today. In 1947, just after Western civilization had narrowly escaped collapse during the horrors of the Second World War, cultural leaders were engaged in intense discussion about how to re-establish civilization on a firmer basis. Religion and science were the old and the new authorities and the crying need of the hour seemed to be to reconcile the two. “Redeeming culture” is an apt title for the enterprise.

In the United States, science was at the height of its prestige. Yet the atom bomb tarnished the scientific image and the nation was also in the midst of a religious revival. How could an adequate moral basis be found for rebuilding the civilization? Many intellectuals thought that religion would be necessary for that, perhaps religion updated in the light of science. Others saw science itself as the only hope for establishing universal values.

It is particularly striking from the perspective of half a century later that so many cultural leaders believed that it would be possible to find a foundational set of values for the whole culture, or at least for its mainstream. Today, ‘pluralism’ and ‘diversity’ are the principal motifs in thinking about culture, and the idea that there could be one dominant way of reconciling science and religion seems quaint. Although there still are many intellectual and popular efforts to relate science to religion, most take place within particular religious traditions and few presume to speak for the entire cultural mainstream.

In the era after the Second World War it was still common to talk about reconciling ‘science and religion’ as though these designated essentially only two points of view. In fact, ‘religion’ included many irreconcilable divisions, including many differ-

ences among Christians themselves. Some of Gilbert's episodes illustrate this. He describes the early decades of the American Scientific Affiliation, founded by conservative Protestants in 1941, which eventually tried to carve out territory between creation scientists on the right and liberal Christians and secularists on the left. In the meantime, representatives of the liberal Protestant establishment at Harvard founded the Institute on Religion in an Age of Science, which had wider ambitions to establish a culture-wide scientific religion. Even though both these organizations had grown out of the American Protestant mainstream, they had opposing views of the essence of their Christian heritage.

Gilbert suggests that the various science and religion dialogues help to explain the persistence of religious faith in the United States, as such exchanges helped to create cultural space for religion. That is probably correct. Yet, if that is the significance of the stories, then comparison is called for. Were the US discussions of religion and science after the war much different from those in the United Kingdom, the Netherlands, Germany or Scandinavia? I suspect that they were much the same at the intellectual level, but that the main difference was the much larger role that evangelical Protestant and other forms of popular religion played in the United States.

By the late 1960s, the mainstream science and religion dialogues of the mid-century were swept aside by cultural forces over which such efforts had little influence. Science declined in cultural prestige. Mainstream Protestantism declined even faster. Yet in the United States evangelicalism as well as popular New Age religions were waiting in the wings. These offered so many varieties of ways of reconciling, or not reconciling, science and religion that it is hard to see how that factor could be the critical variable.

Nonetheless, Gilbert does point out that virtually every group has to negotiate between these cultural authorities and he skilfully provides some valuable background for reflecting on the multiform American ways of doing so.

George Marsden is in the Department of History, University of Notre Dame, Notre Dame, Indiana 46556, USA.

A revised and expanded edition of *No Other Gods: On Science and American Social Thought* by Charles E. Rosenberg has just been published. First issued in 1976, this influential book looks at the ways in which social institutions and values have shaped American scientific practice and thought. Rosenberg, a historian of medicine, now extends the discussion to address the current debate about the social construction of knowledge. Johns Hopkins University Press, \$48.50, £40 (hbk), \$16.95, £14 (pbk).

Evidence for human influence on climate from hemispheric temperature relations

Robert K. Kaufmann* & David I. Stern†

* Center for Energy and Environmental Studies, 675 Commonwealth Avenue, Boston University, Boston, Massachusetts 02215, USA

† Centre for Resource and Environmental Studies, Australian National University, Canberra ACT 0200, Australia

Analysis of observational temperature records for the Northern and Southern hemispheres indicates a statistical relationship in which Northern Hemisphere temperature depends on temperature in the Southern Hemisphere. This pattern, which has strengthened over time, can be explained by the climatic effects of anthropogenic trace gases and tropospheric sulphate aerosols. A similar statistical pattern is produced by model simulations of the historical atmosphere.

There is general agreement that the average surface air temperature of the Earth has increased by about 0.6 °C over the past century^{1–3}. But the cause(s) of the temperature increase is (are) uncertain^{4–6}; the increase may be caused by natural mechanisms, by human activity, or by both.

There is increasing evidence that some of the temperature increase can be attributed to activities that increase the atmospheric concentration of greenhouse gases and tropospheric sulphates⁷. Including these gases in atmospheres used in climate models increases these models' ability to simulate the spatial and temporal pattern of the historical temperature record⁸. Further evidence is provided by analysis of the historical temperature record using spectral analysis^{9–11} or single equation regression models^{9,12–15}. Some of these studies find that anthropogenic variables have a significant effect on global temperatures. Nonetheless, many of these results are difficult to interpret from either a statistical and/or a physical standpoint because there are no clear criteria for the selection of variables, and the specification of dynamic structure and/or the time-series properties of the variables have been ignored in deriving inferences. To alleviate these difficulties, we use here standard econometric time-series criteria for model selection, simulate the distributions of test statistics under different assumptions about the time-series properties of the variables, and search for similar results in the output of a coupled general circulation model (CGCM).

The analysis focuses on the spatial pattern of temperature change that may be generated by economic activities (which occur predominantly in the Northern Hemisphere) that increase the concentration of greenhouse gases globally, but which increase the concentration of tropospheric sulphates in the Northern Hemisphere relative to the Southern Hemisphere. This difference may allow us to attribute temperature change between 1865 and 1994 to economic activity by answering three questions: (1) is there a causal order to temperature changes in the Northern and Southern hemispheres; (2) what factors are responsible for the causal order; and (3) are hypotheses about the factors responsible for causal order consistent with the results generated by CGCMs.

Traditional regression or correlation analysis does not indicate whether the estimated relationship is coincidental or whether the 'dependent' variable is meaningfully dependent on changes in the 'independent' variables. This type of dependence can be examined by testing for Granger causality¹⁶. The presence of Granger causality implies the presence of a statistical causal ordering. Granger causality tests are based on the notion of predictability, in particular whether past values of a variable *X* contain statistically meaningful information about current values of variable *Y* that is not contained

in past values of variable *Y* and other relevant information. Should past values of variable *X* contain information about current values of variable *Y* beyond the information contained in the *Y* sequence and the other variables in the information set, variable *X* is said to 'Granger cause' variable *Y*. The detection of Granger causality does not necessarily imply the presence of a physical causal mechanism between the two variables. Furthermore, the detection of Granger causality depends on the information set of conditioning variables. The coefficient estimates may be biased by the omission of relevant variables that are in fact the causal variables.

Here we look for an anthropogenic influence on global temperature by testing the ability of forcing variables such as trace gas concentrations and sulphate aerosols to explain a detected direction of Granger causality from Southern Hemisphere to Northern Hemisphere temperatures in a model that excludes those former variables. We find that the apparent dependence of Northern Hemisphere on Southern Hemisphere temperatures has strengthened over time and that the best explanation for this pattern is the influence of both anthropogenic trace gases and sulphate aerosols in addition to natural stratospheric sulphates and solar irradiance. To validate this result, we examine the temperature data generated by the Hadley Centre climate model (HCCM) when driven by atmospheres that are consistent with, and different from, the historical record. We find that the presence and direction of causal order is consistent with our hypothesis that the spatiotemporal pattern of greenhouse gases and tropospheric sulphates is responsible for the causal order found in the historical record.

South-to-north causal order

We first look for a causal order between the temperature series in a bivariate autoregressive model—equations (1) and (2). This type of model is known as a vector autoregression in the econometrics literature¹⁷. The test for causality is performed in two steps. In the first step, unrestricted and restricted forms of an equation are estimated. In the second step, a test statistic is calculated to test whether the restriction is binding. We illustrate this procedure by outlining the steps used to test the hypothesis that temperature in the Southern Hemisphere does not Granger cause temperature in the Northern Hemisphere. In the first step, we estimate the unrestricted model, which is given by equations (1) and (2) (which constitute model 1):

$$N_t = \alpha_1 + \beta_1 \text{ time} + \sum_{i=1}^s \delta_{1i} N_{t-i} + \sum_{i=1}^s \phi_{1i} S_{t-i} + \epsilon_{1t} \quad (1)$$

$$S_t = \alpha_2 + \beta_2 \text{ time} + \sum_{i=1}^s \delta_{2i} N_{t-i} + \sum_{i=1}^s \phi_{2i} S_{t-i} + \epsilon_{2t} \quad (2)$$

in which N_t is the temperature anomaly in the Northern Hemisphere^{1,18,19}, S_t is the temperature anomaly in the Southern Hemisphere^{1,18,19}, ϵ_t is a normally distributed random error term, and α_p , β_p , ϕ_{ji} and δ_{ji} are regression coefficients. The constant and time trend are included because univariate tests^{20,21} indicate that the temperature data contain a deterministic trend.

The maximum number of lags, s , that we consider is $5 \approx T^{1/3}$, where T is the number of observations. A likelihood ratio test developed by Sims¹⁷ indicates that s can be reduced to 4 without loss of explanatory power. The same result is indicated by the Akaike information criterion²²—a likelihood-based goodness-of-fit criterion. These tests and sensitivity analyses are performed on the longest possible sample period that can support five lags—1865 to 1994. The earliest observation in our sulphur emission series is 1860.

We test for Granger causality by testing the significance of parametric restrictions in equations (1) and (2). We test whether Southern Hemisphere temperatures Granger cause temperature in the Northern Hemisphere by jointly restricting the ϕ_{1i} to zero thus excluding the lagged values of Southern Hemisphere temperatures from equation (1).

The significance of this restriction is evaluated by the test statistic ω :

$$\omega = \frac{(\text{RSS}_r - \text{RSS}_u)/s}{(\text{RSS}_u)/(T - k)} \quad (3)$$

where T is the number of observations, k is the number of regressors in the unrestricted version of equation (1), s corresponds to the number of coefficients restricted to zero, RSS_r is the residual sum of squares from the restricted version of equation (1) and RSS_u is the residual sum of squares from the unrestricted version of equation (1). The test statistic ω is distributed with the F distribution with s and $(T - k)$ degrees of freedom in the numerator and denominator, respectively, under the assumption that the data series are stationary around a deterministic trend. Values of ω that exceed the critical value indicate that the residual sum of squares for the restricted model increases in a manner that is statistically significant (at the relevant level of significance, $p(F)$) relative to the residual sum of squares for the unrestricted model, in which case we reject the hypothesis of no causal order. To test whether Northern Hemisphere temperatures Granger cause Southern Hemisphere temperatures, we jointly restrict the δ_{2i} to zero, thus excluding the lagged values of Northern Hemisphere temperatures from equation (2).

The results indicate that temperature in the Northern Hemisphere does not Granger cause temperature in the Southern Hemisphere ($\omega(4) = 0.42$, $p < 0.79$). On the other hand, the results indicate that temperature in the Southern Hemisphere does Granger cause temperature in the Northern Hemisphere ($\omega(4) = 3.19$, $p < 0.016$).

If the data series are not trend-stationary but instead contain a random-walk component, the ω statistic cannot be evaluated against the F distribution. This type of non-stationarity is referred to in the econometric literature as the presence of a 'unit root'²³. Univariate tests for this type of non-stationarity^{20,21,24} indicate that the temperature anomalies for the Northern and Southern hemispheres are trend-stationary, and therefore the significance of the ω statistic can be evaluated with the standard F distribution.

The conclusion that the temperature data are trend-stationary is contradicted by results generated by preliminary analyses using multivariate techniques^{25,26}. If greenhouse gases and other forcing factors that are known to contain unit roots drive temperature, the temperature series must contain this unit-root signal. The Dickey–Fuller and other unit-root tests tend to reject the null hypothesis too often when the data-generating process is a random walk with noise

and the noise is large relative to the signal^{23,27}. In the multivariate setting, this noise is reduced and the signal is revealed. Even if the temperature data have a unit root, the significance of the test statistic ω can be evaluated with the F distribution if the data are cointegrated. Cointegration implies that the random-walk processes present in the two series are the same stochastic process. Under these conditions, there is a linear combination of the two variables that is stationary and to which the usual distribution theory can be applied.

If the data for temperature have a unit root and are not cointegrated, the presence of Granger causality cannot be tested accurately using the F distribution: evaluating ω against the F distribution would overstate its statistical significance^{28,29}. This bias would cause us to argue for causality when none is present.

To analyse the distribution of the ω statistic if the temperature data are non-stationary and not cointegrated, we generated 1,000 experimental data sets each with 134 observations for temperature in the Northern and Southern Hemisphere. The model for each temperature series is a random walk with drift:

$$S_t = a_1 + S_{t-1} + e_{1t} \quad (4)$$

$$N_t = a_2 + N_{t-1} + e_{2t} \quad (5)$$

The model is calibrated by regressing the observed data on a constant and time trend and using the estimated regression coefficients for the time trend for a_1 and a_2 and the estimated error variances for the variances of e_{1t} and e_{2t} . The test statistic ω in equation (3) is calculated for each data set and these values are ranked in descending order. This ranking is used to evaluate the significance level ($p(\text{unit root})$) of the test statistics using the simulated distribution. The value in position 50, which corresponds to the 0.05 critical value, is 3.01. The 5% critical value for the F distribution with 4 and 120 degrees of freedom, which is 2.44, is found at position 81, which corresponds to a significance level of 0.081.

The conclusion that Southern Hemisphere temperature 'Granger causes' Northern Hemisphere temperature remains, regardless of the presence of a unit root in the temperature data, when we evaluate the ω statistic against the distribution generated by the experimental data sets ($P < 0.04$).

Explanatory variables

By itself, the south-to-north causal order cannot be used to detect the effect of anthropogenic activities because the causal order may be generated by natural and/or anthropogenic mechanisms. To differentiate between these possibilities, we conduct several tests.

The south-to-north causal order might be generated by episodic, short-run, El Niño/Southern Oscillation (ENSO) teleconnections³⁰. We evaluate this mechanism by analysing the causal order of temperature data averaged over periods thought to be longer than the ENSO teleconnections. If ENSO teleconnections generate the south-to-north causal order, the causal order should not be present in data averaged over long periods. The south-to-north causal order is present in temperature data averaged over the previous 5 years ($\omega(3) = 6.21$, $p < 5.95 \times 10^{-4}$), 10 years ($\omega(6) = 3.64$, $p < 0.003$), and 15 years ($\omega(6) = 3.84$, $p < 0.002$).

To explore the effects of other natural and anthropogenic mechanisms, we repeat the tests of Granger causality with models 2–5 that include natural and/or anthropogenic variables (Box 1). Expanding the model is necessary because we know that model 1 is mis-specified—it omits natural and/or anthropogenic variables that may affect temperature. This omitted-variable bias may generate the Granger causality described in the previous section. When the relevant natural and/or anthropogenic variables are included, this information may weaken or eliminate the change in the residual sum of squares that signals Granger causality.

To investigate the effect of natural and anthropogenic factors, we expand equation (1) to include exogenous variables:

$$N_t = \alpha_1 + \beta_1 \text{time} + \sum_{i=1}^s \delta_{1i} N_{t-i} + \sum_{i=1}^s \phi_{1i} S_{t-i} + \sum_{j=1}^k \sum_{i=1}^s \Gamma_{ji} Z_{jt-i} + \epsilon_{1t} \quad (6)$$

Box 1 Models tested

Model 1. Temperature only

$$\text{Temp} = \alpha + \beta \text{time} + \sum_{i=1}^k \delta_i N_{t-i} + \sum_{i=1}^k \phi_i S_{t-i}$$

Model 2. Natural variables

$$\text{Temp} = \alpha + \beta \text{time} + \sum_{i=1}^s \delta_i N_{t-i} + \sum_{i=1}^s \phi_i S_{t-i} + \sum_{i=1}^s \Gamma_{1i} \text{Sun}_{t-i} + \sum_{i=1}^s \Gamma_{2i} \text{SSN}_{t-i} + \sum_{i=1}^s \Gamma_{3i} \text{SSS}_{t-i}$$

Model 3. Greenhouse gases

$$\text{Temp} = \alpha + \beta \text{time} + \sum_{i=1}^s \delta_i N_{t-i} + \sum_{i=1}^s \phi_i S_{t-i} + \sum_{i=1}^s \Gamma_{1i} \text{Sun}_{t-i} + \sum_{i=1}^s \Gamma_{2i} \text{SSN}_{t-i} + \sum_{i=1}^s \Gamma_{3i} \text{SSS}_{t-i} + \sum_{i=1}^s \Gamma_{4i} \text{CO2}_{t-i} + \sum_{i=1}^s \Gamma_{5i} \text{CH4}_{t-i} + \sum_{i=1}^s \Gamma_{6i} (\text{CFC11}_{t-i} + \text{CFC12}_{t-i} + \text{N2O}_{t-i})$$

Model 4. Tropospheric sulphates

$$\text{Temp} = \alpha + \beta \text{time} + \sum_{i=1}^s \delta_i N_{t-i} + \sum_{i=1}^s \phi_i S_{t-i} + \sum_{i=1}^s \Gamma_{1i} \text{Sun}_{t-i} + \sum_{i=1}^s \Gamma_{2i} \text{SSN}_{t-i} + \sum_{i=1}^s \Gamma_{3i} \text{SSS}_{t-i} + \sum_{i=1}^s \Gamma_{7i} \text{SOX}_{t-i}$$

Model 5. Greenhouse gases and tropospheric sulphates

$$\text{Temp} = \alpha + \beta \text{time} + \sum_{i=1}^s \delta_i N_{t-i} + \sum_{i=1}^s \phi_i S_{t-i} + \sum_{i=1}^s \Gamma_{1i} \text{Sun}_{t-i} + \sum_{i=1}^s \Gamma_{2i} \text{SSN}_{t-i} + \sum_{i=1}^s \Gamma_{3i} \text{SSS}_{t-i} + \sum_{i=1}^s \Gamma_{4i} \text{CO2}_{t-i} + \sum_{i=1}^s \Gamma_{5i} \text{CH4}_{t-i} + \sum_{i=1}^s \Gamma_{6i} (\text{CFC11}_{t-i} + \text{CFC12}_{t-i} + \text{N2O}_{t-i}) + \sum_{i=1}^s \Gamma_{7i} \text{SOX}_{t-i}$$

The symbols used above have meanings as follows: CFC11, radiative forcing of CFC11 (W m^{-2} ; refs 40, 41); CFC12, radiative forcing of CFC12 (W m^{-2} ; refs 40, 41); CH4, radiative forcing of methane (W m^{-2} ; refs 42–45); CO2, radiative forcing of carbon dioxide (W m^{-2} ; refs 46, 47); N, temperature anomaly in the Northern Hemisphere ($^{\circ}\text{C}$; refs 1, 18, 19); N2O, radiative forcing of nitrous oxide (W m^{-2} ; refs 48–50); S, temperature anomaly in the Southern Hemisphere ($^{\circ}\text{C}$; refs 1, 18, 19); SOX, radiative forcing due to anthropogenic emissions of sulphates (W m^{-2} ; ref. 51); SSN, radiative forcing due to stratospheric sulphates, Northern Hemisphere (W m^{-2} ; ref. 52); SSS, radiative forcing due to stratospheric sulphates, Southern Hemisphere (W m^{-2} ; ref. 52); Sun, solar activity (W m^{-2} ; ref. 15); Temp, Temperature anomalies Northern or Southern Hemisphere ($^{\circ}\text{C}$; refs 1, 18, 19). Greek characters are regression coefficients. The subscripts associated with equations for the Northern and Southern Hemisphere (equations (1) and (2)) are suppressed.

Radiative forcing is calculated from atmospheric concentration using the equations in Shine *et al.*⁵³ and Kattenberg *et al.*⁵⁴; radiative forcing for CFCs accounts for the effect of ozone depletion, and the radiative forcing for nitrous oxide accounts for overlap with methane.

in which Z_{jt-i} are vectors of lagged values of exogenous variables and Γ_{ji} are regression coefficients. The time-series properties of the exogenous variables may affect the distribution of the test statistic. The distribution is unaffected by the addition of random walks if the temperature data are trend-stationary²⁸. But if the temperature data contain a unit root, the distribution depends on the number of non-stationary variables added to the model. Univariate tests indicate that the radiative forcing of solar activity, anthropogenic emissions of sulphur, and the atmospheric concentrations of carbon dioxide,

Table 1 Tests of Granger causality

	Lags	ω	$p(F)$	p (unit root)
Historical record				
Model 1 $\bar{R}^2 = 0.64416$				
North causes south	4	0.42	0.79	0.843
South causes north	4	3.19	0.016	0.039
Model 2 $\bar{R}^2 = 0.65674$				
North causes south	4	0.74	0.57	0.645
South causes north	4	3.33	0.013	0.027
Model 3 $\bar{R}^2 = 0.68989$				
North causes south	4	0.27	0.14	0.925
South causes north	4	2.14	0.081	0.151
Model 4 $\bar{R}^2 = 0.65211$				
North causes south	4	0.55	0.70	0.764
South causes north	4	3.50	0.010	0.028
Model 5 $\bar{R}^2 = 0.68599$				
North causes south	4	0.37	0.83	0.867
South causes north	4	1.87	0.12	0.206
HCCM experiments				
GHG 1861–1990				
North causes south	1	0.22	0.88	0.742
South causes north	1	2.74	0.10	0.247
SUL 1861–1990				
North causes south	1	0.01	0.91	0.932
South causes north	1	3.63	0.06	0.172
GHG 1991–2099				
North causes south	1	4.70	0.03	0.115
South causes north	1	1.75	0.19	0.353
SUL 1991–2099				
North causes south	1	5.03	0.02	0.093
South causes north	1	1.38	0.24	0.407

ω is as defined in equation (3), $p(F)$ is the significance level using the F distribution, and p (unit root) is the significance level using our simulated unit root distribution. In the 'historical record' data above, $\bar{R}^2$ indicates the adjusted R^2 value; two alternative hypotheses are tested for each model, 1–5. HCCM indicates Hadley Centre climate model (see text). Values that exceed the 0.05 threshold for statistical significance are shown in bold.

Table 2 Sensitivity analysis

Test no.	Change relative to base models				Model 1		Model 5	
	1864	Aggregate GHGS	5 lags	Contemporaneous effects	ω	$p(F)$	ω	$p(F)$
1	✓				2.44	0.011	2.31	0.063
2		✓			3.19	0.016	2.40	0.055
3			✓		2.61	0.028	0.95	0.45
4				✓	3.19	0.016	1.37	0.24
5	✓	✓			3.44	0.011	2.95	0.024
6	✓			✓	3.44	0.011	2.12	0.085
7		✓	✓		2.61	0.028	1.84	0.11
8		✓		✓	3.19	0.016	2.68	0.036
9			✓	✓	2.61	0.028	1.15	0.34
10		✓	✓	✓	2.61	0.028	2.37	0.046
11	✓	✓		✓	3.44	0.011	3.18	0.017

Values that exceed the 0.05 threshold for statistical significance are shown in bold. Tick marks indicate change relative to base models.

methane, CFC11 (CCl_3F), CFC12 (CCl_2F_2) and nitrous oxide contain a unit root. We repeat the Monte Carlo simulations to estimate the distribution of the test statistic for models 2–5 in which the number of random walks is equal to the number of non-stationary variables in the Z vector. As before, the critical levels of the test statistic are increased.

The results of model 2 indicate that solar activity and the radiative

forcing due to stratospheric sulphates do not affect the direction of causality indicated by model 1. Lagged values of temperature in the Southern Hemisphere contain statistically significant information about current values of temperature in the Northern Hemisphere ($\omega(4) = 3.33, p < 0.013$). This result also is significant when tested against the distribution generated with one random walk ($p < 0.027$).

Next we expand the vector of exogenous variables, Z , to also include the effects of anthropogenic activity. The results indicate that the south-to-north causal order is partially eliminated when the effect of greenhouse gases alone is included in model 3 (Table 1). The test statistic is significant at the 10% level but not at the 5% level. The null hypothesis cannot be rejected at the 10% significance level when tested against the distribution generated with four random walks ($p < 0.151$). Including the effect of sulphur emissions alone (model 4) does not reduce the statistical significance of the causal order in temperature (Table 1).

The results of model 5 indicate that the south-to-north causal order may be caused by both greenhouse gases and tropospheric sulphates. When these variables are included, the test statistic is insignificant as evaluated against a distribution generated with five random walks ($p < 0.206$) and the standard F distribution ($p < 0.12$).

Based on the lack of causal order in model 5 and the somewhat lesser reduction in the significance level in model 3, we hypothesize that the south-to-north causal order is generated by anthropogenic activities that increase the concentration of greenhouse gases globally, but which increase the concentration and effects of sulphate aerosols mainly in the Northern Hemisphere. We hypothesize that anthropogenic emissions of carbon dioxide, methane, CFCs and nitrous oxide increase radiative forcing globally because of their long residence time in the atmosphere. This tends to cause the temperature of the Earth to rise. This rise is retarded in the Northern Hemisphere by the presence of tropospheric sulphates. These aerosols spend a relatively short time in the atmosphere³¹, and so their cooling effects are localized in the Northern Hemisphere.

Calculations indicate that the cooling effect of tropospheric sulphates in the Northern Hemisphere is larger than the heating effect of the higher concentration of greenhouse gases in the Northern Hemisphere relative to the Southern Hemisphere. One estimate for the total (both natural and anthropogenic) direct and indirect effect of sulphate aerosols in the Northern Hemisphere (-0.72 W m^{-2}) is about two times greater than the negative effect of sulphate aerosols in the Southern Hemisphere (-0.38 W m^{-2})³². This 0.34 W m^{-2} difference probably is much larger than the small difference in the radiative forcing of greenhouse gases in the Northern and Southern hemisphere, which is estimated to be about 2.1 W m^{-2} globally. For example, the atmospheric concentration of carbon dioxide, CFC11 and CFC12 over the Northern Hemisphere is about 2–3% higher than over the Southern Hemisphere^{33–36}.

The hypothesis that the south-to-north causal order is generated by the spatiotemporal pattern of anthropogenic emissions of trace gases and sulphate aerosols is consistent with changes in the strength of the causal order over time. An iterative search indicates that the south-to-north causal order generally is not present in sample periods that start in 1865 and end before the 1970s, but is present in samples that start in 1865 and end later than the mid-1970s (Fig. 1a). This implies that anthropogenic variables may be responsible for the apparent direction of causality, though some sources of natural variability such as solar irradiance also have increased over this period¹⁵. The effect of anthropogenic activities is consistent with a simple model that indicates the significance level of the causal order is related to solar irradiance and the radiative forcing from greenhouse gases, and the radiative forcing associated with anthropogenic sulphur emissions (Fig. 1a). Standard tests such as the augmented Dickey–Fuller (-4.26) and the cointegrating Durbin–Watson (0.50) indicate that residuals from this

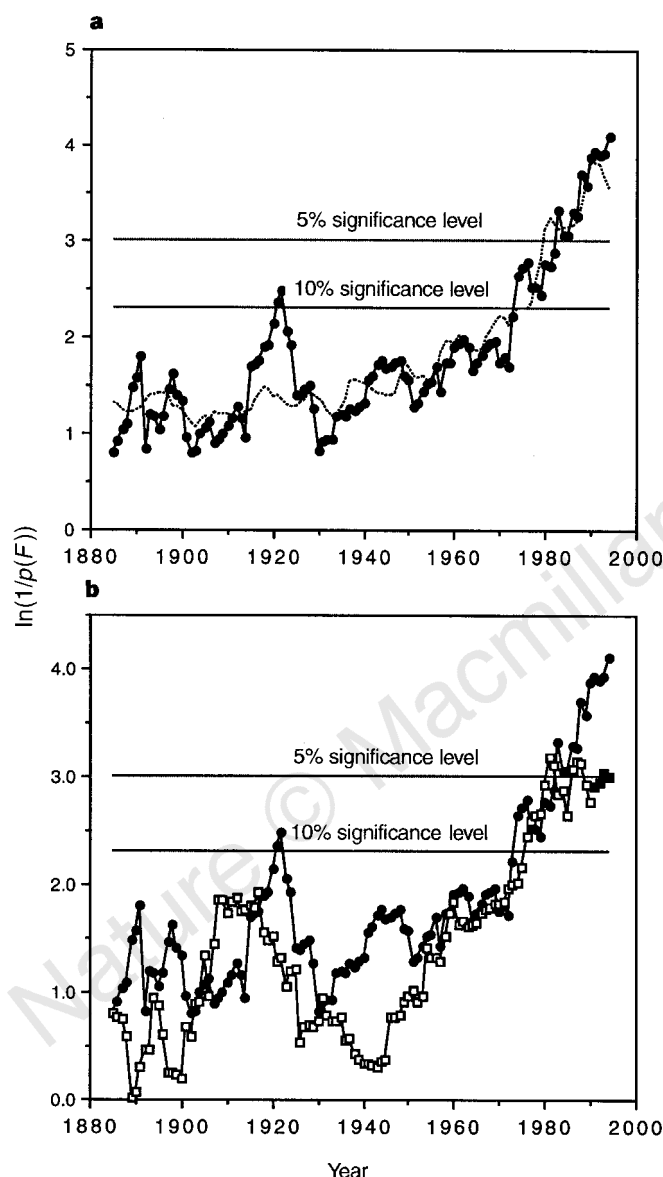


Figure 1 **a**, Changes in the statistical significance of the south-to-north causal order over time (filled circles). This is represented by the natural logarithm of the reciprocal of the significance level for tests carried out on subsamples ending at the dates indicated. The natural logarithm is used because the significance level is bounded from below by zero. The dotted line is the predicted value for $\ln(1/p(F))$ based on the following equation that is fitted using ordinary least squares: $\ln(1/p(F)) = 1.36 - (1.07\text{RFSOX}) + (1.894\text{ORF}) - 0.026 \text{ year}$, where RFSOX is the radiative forcing of sulphates, ORF is the radiative forcing associated with greenhouse gases (CO_2 , CH_4 , CFCs and N_2O) and solar radiation, and 'year' is indexed to one in 1860. **b**, The filled circles are the same variable as in **a**. Open squares are $\ln(1/p(F))$ of the same test on the data from the Hadley Centre SUL experiment (see text) simulated using historical estimates of sulphate aerosol for subsamples ending at the dates indicated. The filled squares are the same test extended to the end of 1994 using IPCC scenarios for greenhouse gases and sulphate aerosols.

regression are stationary, which implies that the variables cointegrate and these two variables adequately model the non-stationarity in the significance level³⁷.

Sensitivity analysis

The relation between the south-to-north causal order and anthropogenic emissions of trace gases and sulphate aerosols also is consistent with the results of a sensitivity analysis in which model 1 and model 5 are estimated using alternative specifications. The alternative specifications include: (1) lengthening the sample period to include observations from 1864; (2) restricting the coefficients associated with the radiative forcing of each of the trace gases to be equal by aggregating their values; (3) increasing the number of lags to 5; and (4) including contemporaneous values of the explanatory variables in model 5. In addition, we examine combinations of these alternative specifications, which generates 11 possible tests (Table 2).

The sensitivity analysis confirms our general result—there is a statistically significant south-to-north causal order in model 1 that is reduced when information about the radiative forcing of trace gases and anthropogenic sulphate emissions are included in model 5. The south-to-north causal order of model 1 is present in all alternative specifications. The reduction in the statistical significance of the test in model 5 versus model 1 is reproduced in all specifications. Two of the alternative specifications, using five lags (test 3) and including contemporaneous effects (test 4) amplify our results by reducing greatly the significance level of the test statistic in model 5.

Cases in which the statistical significance of the south-to-north causal order remains above 5% in model 5 generally are associated with combinations of changes that aggregate the radiative forcing of greenhouse gases and extend the sample period to include observations from 1864. These alternative specifications are more restrictive than those used to generate the results in Table 1. Restricting the coefficients associated with the radiative forcing of individual gases to be equal is problematic, because there is some uncertainty about past levels of these gases and about the radiative forcing formulae, particularly in the case of ozone depletion due to CFCs; there are also differences between the effect on temperature of CO₂ and the other gases, due to possible effects of CO₂ on vegetation. The reliability of both the temperature and radiative-forcing data also declines as the sample is extended into the past.

Hadley Centre climate model

To test further the hypothesis that the south-to-north causal order is generated by anthropogenic emissions of greenhouse gases and tropospheric sulphates, we analyse temperature data generated by a coupled atmosphere–ocean general circulation model. If our hypothesis is correct, two results should emerge: (1) experiments run with atmospheres that emulate the historical concentration of greenhouse gases and tropospheric sulphates should reproduce the south-to-north causal order; and (2) a north-to-south or no causal order may be present when the CGCMs are simulated with atmospheres that have temporal and/or spatial concentrations of greenhouse gases and/or tropospheric sulphates that differ from the historical record.

We analyse results generated by the Hadley Centre climate model (HCCM) because it simulates three experiments that vary the spatial and temporal pattern of all greenhouse gases and tropospheric sulphates; the control experiment, the greenhouse gas (GHG) experiment, and the SUL experiment³⁸. The control experiment holds constant the radiative forcing associated with greenhouse gases and tropospheric sulphates at a level consistent with their pre-industrial concentrations. The GHG experiment allows the radiative forcing of the atmosphere to increase at a rate consistent with the CO₂ equivalent of all greenhouse gases over the 1860–1990 historical period. The SUL experiment adds the effect of tropospheric sulphates to the effect of greenhouse gases.

The concentration of these gases in the 1991–2099 forecast horizon for the GHG and SUL experiments is derived from IPCC emission scenarios. The simulated atmospheres forecast a rapid increase in total radiative forcing associated with greenhouse gases, and a shift south towards the Equator in the spatial distribution of tropospheric sulphates.

The causal order of temperature data for the Northern and Southern Hemisphere found in the control, GHG and SUL experiments over the historical reconstructions and forecast horizons are consistent with the hypothesis that the south-to-north causal order in the historical record is generated by the atmospheric concentration of greenhouse gases and tropospheric sulphates. When the HCCM is simulated with an atmosphere that emulates radiative forcing associated with all greenhouse gases and tropospheric sulphates between 1860 and 1990 (the SUL experiment), applying the causality test to the simulation output suggests ($p < 0.06$) south-to-north causal order (Table 1). An iterative search shows that the strength of the causal order in the SUL experiment follows a pattern similar to the historical record—it generally strengthens over time (Fig. 1b). If the SUL experiment were re-run with historical data to the end of 1994, it may increase the significance level of the causal order in the SUL experiment in much the same way as extending the sample period for 1991 to 1994 increases the strength of the causal order in the historical record. Indeed, using the data for 1991–94 from the IPCC scenarios does increase the strength of the relation beyond the 0.05 threshold ($\omega(1) = 3.93, p < 0.049$). Nonetheless, the strength of the south-to-north causal order is weaker than that found in the historical record, and is less robust to changes in lag length than the results generated by the historical record. This is as would be expected, because the experiment did not simulate changes in variables such as solar activity and ozone depletion, and the model cannot be expected to simulate all aspects of the climate system accurately.

On the other hand, there is no causal order in historical temperature data when the HCCM is simulated with an atmosphere that allows radiative forcing to increase at a rate consistent with the CO₂ equivalent of all greenhouse gases over the 1860–1990 historical period but ignores changes in sulphate aerosols (GHG experiment). The results for the control scenario are uncertain. The dates in the control experiment have no connection to historical events, so we test all possible 130-year samples in the 1850–2099 period for which data are available from the control experiment against a distribution generated by the procedure described by Christiano³⁹. Some periods show a south-to-north causal order if the data are stationary. Subsamples with the largest test statistics are significant at the 1% level. No periods show a causal order if the data are nonstationary as implied by univariate tests of the temperature difference between Northern and Southern Hemisphere: the largest test statistic is significant at the 14.1% level (that is, $p < 0.141$).

The temperature data show a strong north-to-south order over the 1991–2099 period when the HCCM is simulated with an atmosphere that is forced by emission of greenhouse gases and tropospheric sulphates given by the IPCC scenario (Table 1). The north-to-south causal order also is found when the model is simulated with an atmosphere forced only by the emission of greenhouse gases given by the IPCC scenario (Table 1).

We find that a south-to-north causal order is suggested by the data simulated by an atmosphere intended to reproduce the historical record, whereas it is not present or runs in the opposite direction in data from counterfactual simulations. This is consistent with the hypothesis that the south-to-north causal order is generated by the historical combination of greenhouse gases and tropospheric sulphates.

Conclusion

This analysis suggests that human activity has played a role in the historical record of temperature. The south-to-north causal order in

the historical temperature record may be a fingerprint of the spatial and temporal pattern of anthropogenic activities that emitted greenhouse gases and tropospheric sulphates between 1865 and 1994. Combined with other results, our results provide further evidence for the conclusion that “the observed trend in global mean temperature over the past 100 years is unlikely to be entirely natural in origin.”⁷ □

Received 27 August 1996; accepted 29 April 1997.

1. Jones, P. D. Hemispheric surface air temperature variations: a reanalysis and an update to 1993. *J. Clim.* **7**, 1794–1802 (1994).
2. Vinnikov, K. Y., Groisman, P. Y. & Lugina, K. M. Empirical data on contemporary global climate changes (temperature and precipitation). *J. Clim.* **3**, 662–677 (1990).
3. Hansen, J. & Lebedeff, S. Global trends of measured air temperature. *J. Geophys. Res.* **92**, 13345–13372 (1987).
4. Kerr, R. A. Is the world warming or not. *Science* **267**, 612 (1995).
5. Schneider, S. H. Detecting climatic change signals: are there any fingerprints? *Science* **263**, 341–347 (1994).
6. Lindzen, R. S. Some coolness concerning global warming. *Bull. Am. Meteorol. Soc.* **71**, 288–299 (1990).
7. Santer, B. D., Wigley, T. M. L., Barnett, T. P. & Anyamba, E. in *Climate Change 1995: The Science of Climate Change* (eds Houghton, J. T. et al.) 407–433 (Cambridge Univ. Press, 1996).
8. Santer, B. D. et al. A search for human influences on the thermal structure of the atmosphere. *Nature* **382**, 39–46 (1996).
9. Lane, L. J., Nichols, M. H. & Osborn, H. B. Time series analysis of global change data. *Environ. Pollut.* **83**, 63–88 (1994).
10. Kuo, C., Lindberg, C. & Thomson, D. J. Coherence established between atmospheric carbon dioxide and global temperature. *Nature* **343**, 709–714 (1990).
11. Thomson, D. J. The seasons, global temperature, and precision. *Science* **268**, 59–68 (1995).
12. Schönwiese, C. D. Analysis and prediction of global climate temperature change based on multifactor observational statistics. *Environ. Pollut.* **83**, 149–154 (1994).
13. Tol, R. S. J. & de Vos, A. F. Greenhouse statistics—time series analysis. *Theor. Appl. Climatol.* **48**, 63–74 (1993).
14. Tol, R. S. J. Greenhouse statistics—time series analysis: Part II. *Theor. Appl. Climatol.* **49**, 91–102 (1994).
15. Lean, J. L., Beer, J. & Bradley, R. Reconstruction of solar irradiance since 1610: implications for climate change. *Geophys. Res. Lett.* **22**, 3195–3198 (1995).
16. Granger, C. W. J. Investigating causal relations by econometric models and cross spectral models. *Econometrica* **37**, 424–438 (1969).
17. Sims, C. Macroeconomics and reality. *Econometrica* **48**, 1–49 (1980).
18. Nicholls, N. et al. in *Climate Change 1995: The Science of Climate Change* (eds Houghton, J. T. et al.) 138–192 (Cambridge Univ. Press, 1996).
19. Parker, D. E., Jones, P. D., Folland, C. K. & Bevan, A. 1994. Interdecadal changes of surface temperatures since the late 19th century. *J. Geophys. Res.* **99**, 14373–14399 (1994).
20. Dickey, D. A. & Fuller, W. A. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Statist. Assoc.* **74**, 427–431 (1979).
21. Phillips, P. & Perron, P. Testing for a unit root in time series regression. *Biometrika* **75**, 335–346 (1988).
22. Akaike, H. in *2nd Int. Symp. on Information Theory* (eds Petrov, P. N. & Csaki, F.) 267–281 (Akademai Kiado, Budapest, 1973).
23. Hamilton, J. D. *Time Series Analysis* (Princeton Univ. Press, 1994).
24. Schmidt, P. & Phillips, P. C. B. LM tests for a unit root in the presence of deterministic trends. *Oxford Bull. Econ. Statist.* **34**, 357–287 (1992).
25. Johansen, S. Statistical analysis of cointegration vectors. *J. Econ. Dynam. Control* **12**, 231–254 (1988).
26. Johansen, S. & Juselius, K. Maximum likelihood estimation and inference on cointegration with application to the demand for money. *Oxford Bull. Econ. Statist.* **52**, 169–209 (1990).
27. Enders, W. *Applied Econometric Time Series* (Wiley, New York, 1995).
28. Toda, H. Y. & Phillips, P. C. B. The spurious effects of unit roots on exogeneity tests in vector autoregressions: an analytic study. *J. Econometr.* **59**, 229–255 (1993).
29. Ohanian, L. E. The spurious effects of unit roots on vector autoregressions. *J. Econometr.* **39**, 251–266 (1988).
30. Philander, S. G. *El Niño, La Niña, and the Southern Oscillation* (Academic, San Diego, 1990).
31. Alkezweeny, A. J. Trend analyses of sulfur dioxide emissions and sulfate concentrations and their application for global cooling. *Atmósfera* **8**, 91–97 (1995).
32. Kiehl, J. T. & Briegleb, B. P. The relative roles of sulphate aerosols and greenhouse gases in climate forcing. *Science* **260**, 311–314 (1993).
33. Keeling, C. D. et al. in *Aspects of Climate Variability in the Pacific and Western Americas* (ed. Petersen, D. H.) 165–236 (Geophys. Monogr. 55, Am. Geophys. Un., Washington DC, 1989).
34. Siegenthaler, U. & Sarmiento, J. L. Atmospheric carbon dioxide and the ocean. *Nature* **365**, 119–125 (1993).
35. Faser, P. & Derek, N. in *Baseline 91* (eds Dick, C. & Gras, J.) 70–81 (Bureau of Meteorology—CSIRO, Melbourne, 1994).
36. Fraser, P., Penkett, S., Gunson, M., Weiss, R. & Rowland, F. S. in *Concentrations, Lifetimes, and Trends of CFC's Halons, and Related Species* (eds Kay, J., Penket, S. & Ormand, F.) 1.1–1.68 (Rep. No. 1339, NASA, Washington DC, 1994).
37. Engle, R. E. & Granger, C. W. J. Cointegration and error-correction: representation, estimation, and testing. *Econometrica* **55**, 251–276 (1987).
38. Mitchell, J. F. B., Johns, T. C., Gregory, J. M. & Tett, S. F. B. Climate response to increasing levels of greenhouse gases and sulphate aerosols. *Nature* **376**, 501–504 (1995).
39. Christiano, L. J. Searching for a break in GNP. *J. Bus. Econ. Statist.* **10**, 237–250 (1992).
40. Prather, M., McElroy, M., Wofsy, S., Russel, G. & Rind, D. Chemistry of the global troposphere: fluorocarbons as tracers of air motion. *J. Geophys. Res.* **92**, 6579–6613 (1987).
41. Elkins, J. W. et al. in *Trends '93: A Compendium of Data on Global Change* (eds Boden, T. A., Kaiser, D. P., Sepanski, R. J. & Stoss, F. S.) 501–504 (ORNL/CDIAC-65, Carbon Dioxide Information Analysis Center, Oak Ridge Natl Lab., Oak Ridge, TN, 1994).
42. Etheridge, D. M., Pearman, G. I. & Fraser, P. J. in *Trends '93: A Compendium of Data on Global Change* (eds Boden, T. A., Kaiser, D. P., Sepanski, R. J. & Stoss, F. S.) 256–260 (ORNL/CDIAC-65, Carbon Dioxide Information Analysis Center, Oak Ridge Natl Lab., Oak Ridge, TN, 1994).
43. Khalil, M. A. K. & Rasmussen, R. A. Global emissions of methane during the last several centuries. *Chemosphere* **29**, 833–842 (1994).
44. Dlugokenchy, E. J., Lang, P. M., Masarie, K. A. & Steele, L. P. in *Trends '93: A Compendium of Data on Global Change* (ORNL/CDIAC-65, Carbon Dioxide Information Analysis Center, Oak Ridge Natl Lab., Oak Ridge, TN, 1994).
45. Prinn, R. G. et al. *CDIAC World Data Center Dataset No. DB-1001* (Available by anonymous ftp from cdiac.esd@ornl.gov).
46. Etheridge, D. M. et al. Natural and anthropogenic changes in atmospheric CO₂ over the last 1,000 years from air in Antarctic ice and firn. *J. Geophys. Res.* **101**, 4115–4128 (1996).
47. Keeling, C. D. & Whorf, T. P. in *Trends '93: A Compendium of Data on Global Change* (eds Boden, T. A., Kaiser, D. P., Sepanski, R. J. & Stoss, F. S.) 16–26 (ORNL/CDIAC-65, Carbon Dioxide Information Analysis Center, Oak Ridge Natl Lab., Oak Ridge, TN, 1994).
48. Prinn, R. D. et al. Atmospheric emissions and trends of nitrous oxide deduced from ten years of ALE/GAGE data. *J. Geophys. Res.* **95**, 18369–18385 (1990).
49. Prinn, R. D. et al. *CDIAC World Data Center Dataset No. DB-1001* (Oak Ridge, 1995).
50. Machida, T., Nakazawa, T., Fujii, Y., Aoki, A. & Watanabe, O. Increase in the atmospheric nitrous oxide concentration during the last 250 years. *Geophys. Res. Lett.* **22**, 2921–2924 (1995).
51. Stern, D. I. & Kaufmann, R. K. *Estimates of Global Anthropogenic Sulfate Emissions 1860–1993* (Working Paper 9602, Centre for Energy & Environmental Studies, Boston University, 1996); (available at http://cres.anu.edu.au/~dstern/CEES_WP9602.w).
52. Sato, M., Hansen, J. E., McCormick, D. J. & Pollack, J. B. Stratospheric aerosol optical depths, 1850–1990. *J. Geophys. Res.* **98**, 22987–22994 (1993).
53. Shine, K. P. R. G., Derwent, D. J., Wuebbles, D. J. & Mockett, J. J. in *Climate Change: The IPCC Scientific Assessment* (eds Houghton, J. T., Jenkins, G. J. & Ephraim, J. J.) 47–68 (Cambridge Univ. Press, 1991).
54. Kattenberg, A. F. et al. in *Climate Change 1995: The Science of Climate Change* (eds Houghton, J. T. et al.) 289–357 (Cambridge Univ. Press, 1996).

Acknowledgements. We thank the following individuals for data: T. Boden, L. D. Harvey, J. Lean, M. Sato and D. Viner. We thank B. P. Briegleb, M. Friedl, S. Hall, G. Hegerl, J. Key, P. Mayewski, B. Moore III, G. Salvucci, H. Y. Toda, K. Trenberth and P. Young for comments on preliminary versions of this Article.

Correspondence and requests for materials should be addressed to R.K.K. (e-mail: kaufmann@bu.edu).

Detection of ozone on Saturn's satellites Rhea and Dione

K. S. Noll^{*}, T. L. Roush^{†‡}, D. P. Cruikshank[†],
R. E. Johnson[§] & Y. J. Pendleton[†]

^{*} Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA

[†] NASA Ames Research Center, Moffett Field, California 94035, USA

[‡] Department of Geosciences, San Francisco State University, San Francisco, California 94132, USA

[§] University of Virginia, Engineering Physics, Charlottesville, Virginia 22901, USA

The satellites Rhea and Dione orbit within the magnetosphere of Saturn, where they are exposed to particle irradiation from trapped ions. A similar situation applies to the galilean moons Europa, Ganymede and Callisto, which reside within Jupiter's radiation belts. All of these satellites have surfaces rich in water ice^{1,2}. Laboratory studies of the interaction of charged-particle radiation with water ice predicted³ the tenuous oxygen atmospheres recently found on Europa⁴ and Ganymede⁵. However, theoretical investigations did not anticipate the trapping of significantly larger quantities of O₂ within the surface ice⁶. The accumulation of detectable abundances of O₃, produced by the action of ultraviolet or charged-particle radiation on O₂, was also not predicted before being observed on Ganymede⁷. Here we report the identification of O₃ in spectra of the saturnian satellites Rhea and Dione. The presence of trapped O₃ is thus no longer unique to Ganymede, suggesting that special circumstances may not be required for its production.

The observations of Rhea, Dione and Iapetus reported here (Fig. 1) were obtained with the Hubble Space Telescope's Faint Object Spectrograph (Table 1). The data were reduced using the standard procedures maintained by the Space Telescope Science Institute. Geometric albedos were calculated by dividing out the solar spectrum⁸ and by correcting to zero degrees phase angle^{9,10}.

The spectra of both hemispheres of Rhea and Dione show clear minima at 260 ± 5 nm. The Iapetus spectra have shallow minima at 270 ± 10 nm (trailing hemisphere) and 255 ± 10 nm (leading hemisphere). Of the products that can be produced in H₂O ice either by radiolysis or photolysis, only O₃ has an absorption feature near 260 nm (refs 11, 12). The wavelength of the albedo minimum is an excellent match to the wavelength of the feature observed on Ganymede's trailing hemisphere⁷. The inferred O₃/O₂ ratio on Ganymede is close to that expected for equilibrium gas trapped in voids or defects in the ice^{7,13}, adding confidence in the identification of the 260-nm absorber as O₃. Due to the similar surfaces and charged particle fluxes on Rhea and Dione, we identify the Hartley continuum of O₃ with the 260-nm absorption on these satellites. The weakness or absence of the feature on Iapetus is consistent with the lower charged-particle density in the solar wind at Iapetus compared to that found in Saturn's magnetosphere at the radii of Rhea and Dione's orbits. Smaller amounts of O₂ and O₃ are therefore expected in the ice on Iapetus.

To estimate the amount of O₃ in the observed spectra of Rhea, Dione and Iapetus we produced a series of models which tested a range of assumptions concerning the continuum. We find the column abundance of ozone to be $N(\text{O}_3) = (1-6) \times 10^{16} \text{ cm}^{-2}$ for Rhea's leading hemisphere with comparable values for the trailing hemisphere and for both hemispheres of Dione. For Iapetus's icy trailing hemisphere we find $N(\text{O}_3) < 0.5 \times 10^{16} \text{ cm}^{-2}$ with a poor match to the wavelength of the spectral minimum. Absorption coefficients of ozone were calculated from a gas-phase spectrum which was shifted by +6 nm (ref. 14) and broadened to a

Table 1 HST observations of Saturn's satellites

Satellite	Date (UT)	λ (nm)	θ (deg)	α (deg)
Dione (T)	21 Dec. 96	220-480	264	5.9
(L)	28 Nov. 96	220-480	89	5.4
Rhea (T)	23 Oct. 94	190-330	288	4.7
(L)	27 Nov. 96	220-480	95	5.4
Iapetus (T)	22 Oct. 94	190-330	269	4.7
(L)	03 Jan. 96	190-330	91	5.5

λ , wavelength; θ , orbital longitude; α , phase angle; T, trailing hemisphere; L, leading hemisphere.

full-width at half-maximum intensity of 80 nm (ref. 15). Though these effects are observed in the laboratory, there are no measurements of O₃ absorption under conditions similar to those found on the satellites of Saturn and Jupiter, and this remains a continuing source of uncertainty in our estimates of O₃ abundances. However, unless the Hartley continuum of O₃ is significantly broader at 100 K than it is at 77 K, O₃ cannot account for the bulk of the ultraviolet absorption; some other absorber is required. The model shown in Fig. 2 uses a continuum that includes absorption by small amounts of organic residue created by irradiation of water ice and ethane¹⁶ and an additional, linear slope at wavelengths less than 477 nm. We have assumed that the organic residue produced in the laboratory is similar to what might be produced in any icy surface with small quantities of carbon-containing material. The additional slope is needed to match the continued red slope observed at wavelengths <220 nm in the spectrum of Rhea's trailing hemisphere and may be related to the observation that spectral slopes in this part of the spectrum for residues created by irradiation of ice mixtures are dependent on radiation history¹⁷. With this continuum we find $N(\text{O}_3) = 2 \times 10^{16} \text{ cm}^{-2}$. The uncertainties in this estimate are at least a factor of 3.

Ozone is produced in ice by a chain of events starting with the dissociation of H₂O by high-energy charged-particle impacts. Oxygen and hydrogen produced in this way recombine, producing some O₂ and H₂. The H₂ subsequently escapes from the satellite because of its high mobility in an ice lattice and its low mass. Oxygen molecules that remain trapped in the ice are exposed to ultraviolet radiation where an equilibrium between O₂ and O₃ results. If our understanding of the production of O₃ is correct, the presence of O₃ on these satellites implies that O₂ is also present at abundances of the order of 500 times higher than O₃ (refs 7, 13). At sufficiently high density and low temperature¹⁵, pairs of O₂ molecules absorb in a series of weak bands, most easily observed at 580 and 630 nm (ref. 6). The inferred abundance of O₃ on Rhea and Dione is approximately half that found on Ganymede which might lead to O₂ bands with a maximum absorption of 1% of the continuum. Visible-wavelength spectra of the precision needed to identify these weak O₂ bands have not been published for Saturn's satellites, but could be obtained with current technology. The lack of these features would not rule out O₃, but their presence would be strong additional evidence for its presence. Another test would be to search for O₃ on the icy satellites Tethys and Enceladus which orbit Saturn at a distance where the density of charged particles is significantly higher than at Rhea or Dione³. If the total flux of charged particles is the most important factor in the production of O₃, then we expect O₃ to be present on both Tethys and Enceladus. The satellites of Uranus are also ice-rich objects¹⁸, but experience a much lower charged-particle flux than Rhea³ and appear to have carbon contaminants which could compete for oxygen. As a result, we expect a lower abundance of O₃ on these satellites.

The rotation periods of Rhea, Dione and Iapetus are synchronous with their revolution about Saturn so that one hemisphere (leading) always faces in the direction of orbital motion while the opposite (trailing) faces away. Both charged-particle radiation

and micrometeorite impacts can have asymmetric effects on the surfaces, with a higher flux of charged particles striking trailing hemispheres and micrometeorite impacts predominating on the leading hemispheres of the satellites¹⁹. Jupiter's icy satellites, also in synchronous rotation, show many leading–trailing hemisphere dichotomies^{1,19}, most of which are attributable to irradiation of the trailing hemisphere. Saturn's satellites Rhea and Dione exhibit leading–trailing hemisphere differences in colour, albedo²⁰ and ice-band depth², and experience charged-particle fluxes comparable to Ganymede³. Iapetus is marked by an extreme darkening of its leading hemisphere but orbits outside Saturn's magnetosphere where it encounters the lower-density charged-particle environment of the solar wind. Besides changes in albedo, colour and grain properties, Europa, Ganymede and Callisto each have compositional differences between leading and trailing hemispheres. However, the compositional differences defy a single explanation with excesses of SO₂ on Europa's trailing hemisphere^{21,22}, O₂ and O₃ on Ganymede's trailing hemisphere^{6,7}, and SO₂ on Callisto's leading hemisphere²³. All three of Jupiter's icy satellites show evidence of excess oxygen being produced in the surface ices, though it appears that when sulphur is present in the ice it competes effectively for the oxygen and limits the production of O₃. The origin of the sulphur may either be external or may be related to the fraction of non-ice components present on the surfaces of Jupiter's icy satellites^{1,22,23}. Although there is circumstantial evidence linking changes in albedo and spectrophotometric properties of Jupiter's satellites to the observed hemispheric compositional differences, a common causal connection has not been established. Saturn's satellites are

smaller than Jupiter's and the surfaces are nearly pure water ice²; these two significant differences may constrain the conditions under which O₃ is formed and determine if spectrophotometric and compositional changes can be related.

Ozone appears to be present at a similar abundance on both leading and trailing hemispheres of Rhea and Dione. The trailing hemispheres of both satellites have lower albedos than their leading hemispheres, but the difference between the albedo in the ultraviolet absorption feature at 260 nm and the albedo in the visible wavelength continuum at 600 nm is very similar, 0.3–0.35 albedo units. Differences in albedo at visible wavelengths between leading and trailing hemispheres of Rhea and Dione must be related to the presence of small quantities of non-ice materials such as organic residues or minerals that will also absorb at ultraviolet wavelengths. A natural explanation of the apparent similarity in the shape of the feature would arise if the surfaces could be described as areal mixtures of spectrally distinct material. The similarity of the O₃ absorption depths on both hemispheres of Rhea and Dione requires that the charged particles initiating O₂ and O₃ production have energies sufficiently high that their incidence on the surface is nearly isotropic. It has been shown that 20-keV oxygen ions will strike Dione nearly isotropically whereas colder oxygen ions (0.2-keV) do not²⁴, possibly indicating a minimum required energy for obtaining a measurable quantity of O₃. The energy of an ion with a gyroradius equal to the satellite radius is 43 times higher for Ganymede than for Dione³, a fact that may account for the apparent difference in longitudinal distribution of O₃ between Dione and Rhea compared to Ganymede.

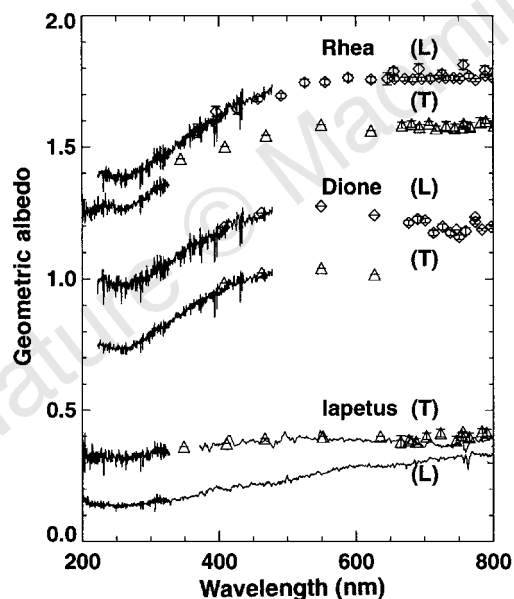


Figure 1 Geometric albedos of Rhea, Dione and Iapetus as a function of wavelength. Spectra of both leading (L) and trailing (T) hemispheres are shown. The thick solid curves are the HST data reported here. Discrete points are from Roush *et al.*¹⁹ from a variety of sources, thin solid curves are scaled reflectances from Vilas *et al.*²⁰. Dione albedos are offset by +0.5, Rhea albedos by +1.0, and the spectrum of Iapetus' dark leading hemisphere has been scaled by a factor of 2.5 for clarity. Spectra of Iapetus and Rhea's trailing hemisphere were corrected for scattered light at wavelengths below 220 nm. The correction is small for wavelengths greater than 200 nm. Both hemispheres of Rhea and Dione have pronounced minima at 260 nm. The Iapetus spectra also appear to have weak minima, but at different wavelengths than Rhea and Dione. For Rhea's trailing hemisphere we also obtained data at wavelengths less than 220 nm. The data here suggest that there may be a second minimum at ~210 nm, though the signal-to-noise ratio is too low to be definitive. A similar minimum was noted in Callisto's spectrum²³, but has not been identified.

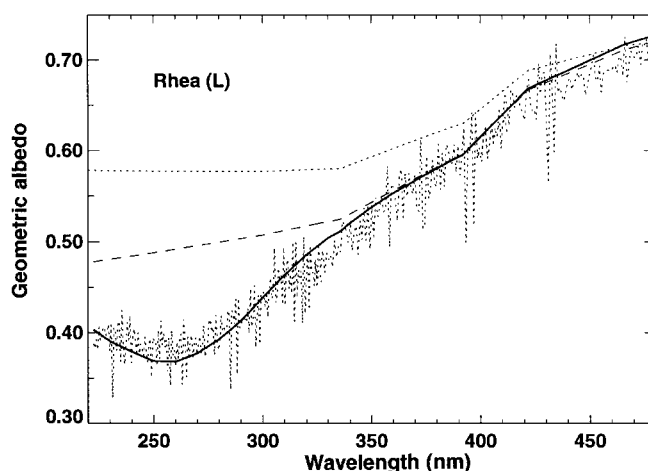


Figure 2 The predictions of a model including O₃ (solid line), plotted with the spectrum of Rhea's leading (L) hemisphere (dotted curve). The continuum consists of a mixture of 97.5% ice and 2.5% organic residues from an irradiated 6:1 mixture of H₂O ice and C₂H₆ ice¹⁶ (sometimes called ice tholins, upper dotted line) with an additional constant-slope absorber below 477 nm (dashed line). This added absorption is required both to fit the observed spectrum from 350 to 450 nm and also to account for the fact that at wavelengths below 220 nm the albedo of Rhea's trailing hemisphere remains low. Other mixes of candidate continuum absorbers are also possible, but are not tested in our models.

The detection of O_3 on Rhea and Dione is the first evidence that the accumulation of oxygen and ozone in icy surfaces exposed to ion irradiation may be a widespread process with significance in at least two areas outside the study of the Solar System. Molecular oxygen has been predicted to be abundant in the mantles of interstellar grains in dense clouds²⁵ and production of O_3 in these grains via irradiation by cosmic rays, stellar winds, or ultraviolet light may be an important factor to consider in models of grain chemistry¹¹. Possibly of greater significance, plans to search for and identify Earth-like planets around other stars frequently cite the 9.8- μm band of O_3 as a means to identify planets with massive O_2 atmospheres²⁶. A significant component of the Earth's atmosphere was probably derived from impacts of icy comets²⁷. If the ices in these comets contain excess oxygen produced by irradiation as the comet accretes, the volatiles delivered to a planetary atmosphere in this way could contain as much as a few per cent O_2 by mass (in excess of the oxygen contained in H_2O ice). Because ozone is a very nonlinear tracer of molecular oxygen in a planetary atmosphere, an Earth-like atmosphere with a partial pressure of O_2 of only 2 mbar, an amount that could conceivably be supplied by O_2 -saturated ices, would have a detectable column of O_3 only one-quarter that in Earth's present atmosphere²⁸ (2 mbar is 2% of Earth's sea-level pressure or $\sim 10\%$ of the O_2 currently in the atmosphere). The identification of oxygen-containing atmospheres on extrasolar planets through detection of O_3 may not, therefore, be an unambiguous way of identifying planets with Earth-like biology. $\square$

Received 27 March; accepted 13 May 1997.

- Clark, R. N. Ganymede, Europa, Callisto, and Saturn's rings: Compositional analysis from reflectance spectroscopy. *Icarus* **44**, 388–409 (1980).
- Clark, R. N., Brown, R. H., Owensby, P. D. & Steele, A. Saturn's satellites: Near-infrared spectrophotometry (0.65–2.5 μm) of the leading and trailing sides and compositional implications. *Icarus* **58**, 265–281 (1984).
- Johnson, R. E. *Energetic Charged Particle Interactions with Atmospheres and Surfaces* (Springer, Berlin, 1990).
- Hall, D. T., Strobel, D. F., Feldman, P. D., McGrath, M. A. & Weaver, H. A. Detection of an oxygen atmosphere on Jupiter's moon Europa. *Nature* **373**, 677–679 (1995).
- Barth, C. A. *et al.* Galileo ultraviolet observations of atomic hydrogen in the atmosphere of Ganymede. *Geophys. Res. Lett.* (in the press).
- Spencer, J. R., Calvin, W. M. & Person, M. J. Charge-coupled device spectra of the galilean satellites: Molecular oxygen on Ganymede. *J. Geophys. Res.* **100**, 19049–19056 (1995).
- Noll, K. S., Johnson, R. E., Lane, A. L., Domingue, D. & Weaver, H. A. Detection of ozone on Ganymede. *Science* **273**, 341–343 (1996).
- Woods, T. N., Rottman, G. J. & Ucker, G. J. Solar-stellar irradiance comparison experiment 1: 2. Instrument calibrations. *J. Geophys. Res.* **98**, 10679–10694 (1993).
- Buratti, B. & Veeverka, J. Voyager photometry of Rhea, Dione, Tethys, Enceladus, and Mimas. *Icarus* **58**, 254–264 (1984).
- Morrison, D., Jones, T. J., Cruikshank, D. P. & Murphy, R. E. The two faces of Iapetus. *Icarus* **24**, 157–171 (1975).
- Elsila, J., Allamandola, L. J. & Sandford, S. A. The 2140 cm^{-1} (4.673 microns) solid CO band: The case for interstellar ice analogs. *Astrophys. J.* (in the press).
- Matich, A. J., Bakker, M. G., Lennon, D., Quickenden, T. I. & Freeman, C. G. O_2 luminescence from UV-excited H_2O and D_2O ices. *J. Phys. Chem.* **97**, 10539–10553 (1993).
- Johnson, R. E. & Jessor, W. A. O_2/O_3 microatmospheres in the surface of Ganymede. *Astrophys. J.* **480**, L79–L82 (1997).
- Taube, H. Photochemical reactions of ozone in solution. *Trans. Faraday Soc.* **53**, 656–665 (1956).
- Vaida, V., Donaldson, D. J., Strickler, S. J., Stephens, S. L. & Birks, J. W. A reinvestigation of the electronic spectra of ozone: Condensed-phase effects. *J. Phys. Chem.* **93**, 506–508 (1989).
- Khare, B. N. *et al.* Production of optical constants of ice tholin from charged particle irradiation of (1:6) C_2H_5/H_2O at 77 K. *Icarus* **103**, 290–300 (1993).
- Thompson, W. R., Murray, B., Khare, B. N. & Sagan, C. Coloration and darkening of methane clathrate and other ices by charged particle irradiation: Application to the outer solar system. *J. Geophys. Res.* **92**, 14933–14947 (1987).
- Roush, T. L., Cruikshank, D. P. & Owen, T. C. in *Volatiles in the Earth and Solar System* (ed Farley, K. A.) 143–153 (Am. Inst. Phys., New York, 1995).
- Buratti, B. Ganymede and Callisto: Surface textural dichotomies and photometric analysis. *Icarus* **92**, 312–323 (1991).
- Buratti, B., Moshier, J. A. & Johnson, T. V. Albedo and color maps of the saturnian satellites. *Icarus* **87**, 339–357 (1990).
- Lane, A. L., Nelson, R. M. & Matson, D. L. Evidence for sulphur implantation in Europa's UV absorption band. *Nature* **292**, 38–39 (1981).
- Noll, K. S., Weaver, H. A. & Gonnella, A. M. The albedo spectrum of Europa from 2200 Å to 3300 Å. *J. Geophys. Res.* **100**, 19057–19059 (1995).
- Noll, K. S., Johnson, R. E., McGrath, M. & Caldwell, J. J. Detection of SO_2 on Callisto with the Hubble Space Telescope. *Geophys. Res. Lett.* **24**, 1139–1142 (1997).
- Pospieszalska, M. K. & Johnson, R. E. Magnetospheric ion bombardment profiles of satellites: Europa and Dione. *Icarus* **78**, 1–13 (1989).
- Tielens, A. G. G. M. & Hagen, W. Model calculations of the molecular composition of interstellar grain mantles. *Astron. Astrophys.* **114**, 245–260 (1982).
- Burke, B. F. Detection of planetary systems and the search for evidence of life. *Nature* **322**, 340–341 (1986).
- Owen, T., Bar-Nun, A. & Kleinfeld, I. Possible cometary origin of heavy noble gases in the atmospheres of Venus, Earth and Mars. *Nature* **358**, 43–46 (1992).

- Léger, A., Pirre, M. & Marceau, F. J. Relevance of oxygen and ozone detections in the search for primitive life in extrasolar planets. *Adv. Space Res.* **14**, 117–122 (1994).
- Vilas, F., Larson, S. M., Stockstill, K. R. & Gaffey, M. J. Unraveling the zebra: Clues to the Iapetus dark material composition. *Icarus* **124**, 262–267 (1996).

Acknowledgements. We thank M. J. Bartholomew for sharing her models of visible and IR satellite spectra on which our models of the UV spectrum are based. This work was supported by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract.

Correspondence should be addressed to K.S.N.

Modelling the evolution of human trail systems

Dirk Helbing*, Joachim Keltsch† & Péter Molnár‡

* Institute of Theoretical Physics, University of Stuttgart, Pfaffenwaldring 57/III, 70550 Stuttgart, Germany

† Science–Computing, Hagellocher Weg 71, 72070 Tübingen, Germany

‡ Center for Theoretical Studies of Physical Systems, Clark Atlanta University, James P. Brawley Drive, Atlanta, Georgia 30314, USA

Many human social phenomena, such as cooperation^{1–3}, the growth of settlements⁴, traffic dynamics^{5–7} and pedestrian movement^{7–10}, appear to be accessible to mathematical descriptions that invoke self-organization^{11,12}. Here we develop a model of pedestrian motion to explore the evolution of trails in urban green spaces such as parks. Our aim is to address such questions as what the topological structures of these trail systems are¹³, and whether optimal path systems can be predicted for urban planning. We use an 'active walker' model^{14–19} that takes into account pedestrian motion and orientation and the concomitant feedbacks with the surrounding environment. Such models have previously been applied to the study of complex structure formation in physical^{14–16}, chemical¹⁷ and biological^{18,19} systems. We find that our model is able to reproduce many of the observed large-scale spatial features of trail systems.

Previous studies have shown that various observed self-organization phenomena in pedestrian crowds can be simulated very realistically. These phenomena include the emergence of lanes of uniform walking direction and oscillatory changes of the passing direction at bottlenecks^{7,10}. Another interesting collective effect of pedestrian motion, which we have investigated very recently, is the formation of trail systems in green areas. In many cases, the



Figure 1 Between the straight, paved ways on the university campus in Stuttgart-Vaihingen a trail system has evolved (centre of the picture). Two types of nodes are observed: intersections of two trails running in a straight line and junctions of two trails which smoothly merge into one trail.

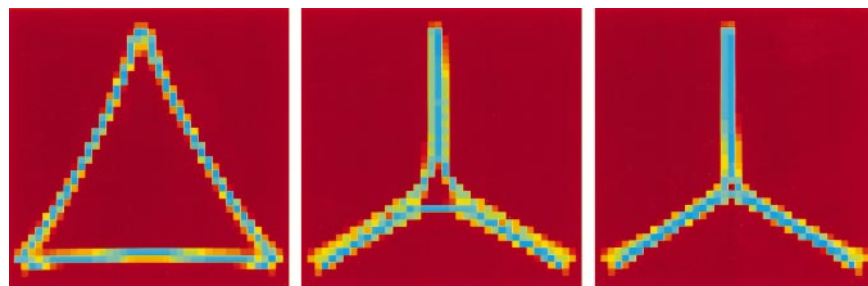


Figure 2 The structure of the emerging trail system (yellow to blue) depends essentially on the attractiveness parameter κ . If κ is small, a direct way system develops (left); if κ is large, a minimal way system is formed (right); otherwise, a compromise between both extremes will result (middle) which looks similar to the trail system in the centre of Fig. 1.

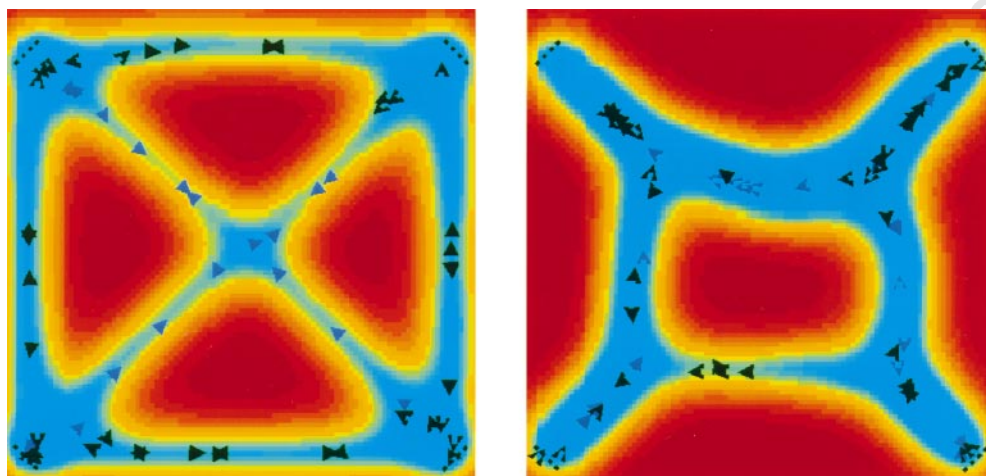


Figure 3 The places and walking directions of pedestrians are represented here by arrows. The trail potential $V_{tr}(\mathbf{r}, t)$ is represented by a colour scale (red indicates small values, blue indicates large values). Starting with a plain ground, the structure of the trail system changes considerably during the simulation. Initially, pedestrians use the direct ways (left). Because frequently used trails become

more comfortable, a bundling of trails sets in which reduces the overall length of the trail system (right). The resulting way system (whose asymmetry is caused by differences in the frequency of trail usage) could serve as a planning guideline. It provides a suitable compromise between minimal construction costs and maximal comfort. Moreover, it balances the relative detours of all walkers.

pedestrians' desire to take the shortest way and the specific properties of the terrain are insufficient for an explanation of the trail characteristics. It is essential to include the effect of human orientation. To simulate the typical features of trail systems, we have extended the earlier model of pedestrian motion to an active-walker model by introducing equations for environmental changes and their effect on the chosen walking direction.

First, we represent the ground structure at place $\mathbf{r}$ and time t by a function $G(\mathbf{r}, t)$ which reflects the comfort of walking. Trails are characterized by particularly large values of G . On the one hand, at their positions $\mathbf{r} = \mathbf{r}_\alpha(t)$, all pedestrians α leave footprints on the ground (for example, by trampling down some vegetation). Their intensity is assumed to be $I(\mathbf{r})[1 - G(\mathbf{r}, t)/G_{\max}(\mathbf{r})]$, as the clarity of a trail is limited to a maximum value $G_{\max}(\mathbf{r})$. This causes a saturation effect $[1 - G(\mathbf{r}, t)/G_{\max}(\mathbf{r})]$ of the ground's alteration by new footprints. On the other hand, the ground structure changes owing to the regeneration of vegetation. This will lead to a restoration of the natural ground conditions $G_0(\mathbf{r})$ with a certain weathering rate $1/T(\mathbf{r})$ which is related to the durability $T(\mathbf{r})$ of trails. Thus, the equation of environmental changes reads:

$$\frac{dG(\mathbf{r}, t)}{dt} = \frac{1}{T(\mathbf{r})} [G_0(\mathbf{r}) - G(\mathbf{r}, t)] + I(\mathbf{r}) \left[1 - \frac{G(\mathbf{r}, t)}{G_{\max}(\mathbf{r})} \right] \sum_{\alpha} \delta(\mathbf{r} - \mathbf{r}_\alpha(t)) \quad (1)$$

where $\delta(\mathbf{r} - \mathbf{r}_\alpha)$ denotes Dirac's delta function (which yields only a contribution for $\mathbf{r} = \mathbf{r}_\alpha$).

The attractiveness of a trail segment at place $\mathbf{r}$ from the perspective of place $\mathbf{r}_\alpha$ decreases with its distance $|\mathbf{r} - \mathbf{r}_\alpha(t)|$ and depends on the visibility $\sigma(\mathbf{r}_\alpha)$. Taking this into account by using a factor $\exp(-|\mathbf{r} - \mathbf{r}_\alpha|/\sigma(\mathbf{r}_\alpha))$ and taking the spatial average by integration of the weighted ground structure over the green area, we obtain:

$$V_{tr}(\mathbf{r}_\alpha, t) = \int d^2r e^{-|\mathbf{r} - \mathbf{r}_\alpha|/\sigma(\mathbf{r}_\alpha)} G(\mathbf{r}, t) \quad (2)$$

The trail potential $V_{tr}(\mathbf{r}_\alpha, t)$ reflects the attractiveness of walking at place $\mathbf{r}_\alpha$. It describes indirect long-range interactions via environmental changes, which are essential for the characteristics of the evolving patterns¹⁸.

On a plain, homogenous ground, the walking direction $\mathbf{e}_\alpha$ of pedestrian α is determined by the direction of the next destination $\mathbf{d}_\alpha$, that is $\mathbf{e}_\alpha(\mathbf{r}_\alpha) = (\mathbf{d}_\alpha - \mathbf{r}_\alpha)/|\mathbf{d}_\alpha - \mathbf{r}_\alpha|$. Without a destination, a pedestrian is expected to move into the direction of the largest increase of ground attraction, which is given by the (normalized) gradient $\nabla_{\mathbf{r}_\alpha} V_{tr}(\mathbf{r}_\alpha, t)$ of the trail potential. However, because the choice of the walking direction $\mathbf{e}_\alpha$ is influenced by the destination and existing trails at the same time, the orientation relation

$$\mathbf{e}_\alpha(\mathbf{r}_\alpha, t) = \frac{\mathbf{d}_\alpha - \mathbf{r}_\alpha + \nabla_{\mathbf{r}_\alpha} V_{tr}(\mathbf{r}_\alpha, t)}{|\mathbf{d}_\alpha - \mathbf{r}_\alpha + \nabla_{\mathbf{r}_\alpha} V_{tr}(\mathbf{r}_\alpha, t)|} \quad (3)$$

was taken as the arithmetic average of both effects. Considering cases of rare interactions, the approximate equation of motion of a

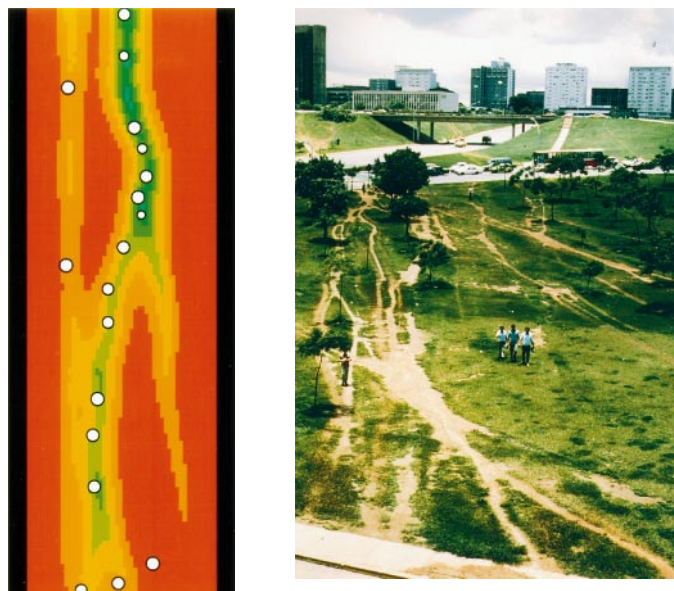


Figure 4 When pedestrians leave footprints on the ground, trails will develop, and only parts of the ground are used for walking (in contrast to paved areas). The similarity between the simulation result (left) and the trail system on the university campus of Brasilia (right, reproduction by permission of K. Humpert) is obvious.

pedestrian α with desired velocity v_α^0 is:

$$\frac{d\mathbf{r}_\alpha}{dt} = v_\alpha^0 \mathbf{e}_\alpha(\mathbf{r}_\alpha, t) \quad (4)$$

A comparison of simulation results with photographs shows that the above described model is in good agreement with empirical observations. In particular, the evolution of the unexpected 'island' in the middle of the trail system in Fig. 1 can be correctly described (Fig. 2). The goodness of fit of the model is quite surprising, as it contains only two independent parameters $\kappa = IT/\sigma$ and $\lambda = V^0 T/\sigma$, where V^0 denotes the average of the desired velocities v_α^0 . This can be shown by scaling the model to dimensionless equations. The parameter λ was kept constant.

Our simulations are based on a discretization of the considered area in small quadratic elements of equal size, which converts the integral in equation (2) into a sum. Temporal and spatial derivatives are approximated by difference quotients. The examples we present here begin with plain, homogenous ground. All pedestrians have their own destinations and entry points, from which they start at a randomly chosen point in time. In Fig. 2 (Fig. 3) pedestrians move between all possible pairs of three (four) fixed places; in Fig. 4, the entry points and destinations are distributed over the small ends of the ground.

At the beginning, pedestrians take the direct ways to their respective destinations. But after some time they begin to use already existing trails, as this is more comfortable than clearing new ways. By this, a kind of selection process^{16,20,21} between trails sets in: frequently used trails are more attractive than others. For this reason they are chosen very often, and the resulting reinforcement makes them even more attractive. However, the weathering effect destroys rarely used trails and limits the maximum length of the way system which can be supported by a certain rate of trail usage. As a consequence, the trails begin to bundle, especially where different trails meet or intersect. This explains why pedestrians with different destinations use and produce common parts of the trail system (Figs 2 and 3).

A direct way system (which provides the shortest connections, but covers a lot of space) only develops if all ways are almost equally comfortable. If the advantage κ of using existing trails is large, the final trail system is a minimal way system (which is the shortest way system that connects all entry points and destinations). For realistic values of κ , the evolution of the trail system stops before this state is reached (Fig. 2). Thus, κ is related to the average relative detour of the walkers. We conjecture (but cannot yet prove) that the resulting

way system is the shortest one which is compatible with a certain accepted relative detour. In this sense, it yields an optimal compromise between convenience and shortness.

Therefore, we suggest the above model could be used by urban planners and landscape gardeners, who face the task of constructing the most comfortable and convenient way systems at minimal construction costs. For planning purposes, one needs to know the entry points and destinations within the considered area and the rates of usage of their connections. If necessary, these can be estimated by trip chaining models²², which are also needed in cases of complex lines of access and sight. The effects of the physical terrain and already existing ways can be taken into account by the function $G_0(\mathbf{r})$. By varying the model parameter κ , the overall length of the resulting trail system can be influenced (Figs 2 and 3). In the same way, one can check its structural stability. We are at present evaluating typical parameter values of κ and λ by comparison of simulation results with real pedestrian flows which are reconstructed from video films by image processing. These values could be used for designing convenient way systems in residential areas, parks and recreation areas by means of computer simulations (Fig. 3).

It would be interesting to relate our model to work on space syntax²³. Repulsive interactions between pedestrians, which are relevant in cases of frequent pedestrian interactions, can be taken into account by generalizing equation (4) in accordance with the social force model of pedestrian motion^{7,10}; in general, they lead to broader trails. $\square$

Received 31 January; accepted 6 May 1997.

1. Axelrod, R. & Dion, D. The further evolution of cooperation. *Science* **242**, 1385–1390 (1988).
2. Clearwater, S. H., Huberman, B. A. & Hogg, T. Cooperative solution of constraint satisfaction problems. *Science* **254**, 1181–1183 (1991).
3. Helbing, D. *Quantitative Sociodynamics Stochastic Methods and Models of Social Interaction Processes* (Kluwer Academic, Dordrecht, 1995).
4. Makse, H. A., Havlin, S. & Stanley, H. E. Modelling urban growth patterns. *Nature* **377**, 608–612 (1995).
5. Herman, R., Lam, T. & Prigogine, I. Multilane vehicular traffic and adaptive human behavior. *Science* **179**, 918–920 (1973).
6. Herman, R. & Prigogine, I. A two-fluid approach to town traffic. *Science* **204**, 148–151 (1979).
7. Helbing, D. *Verkehrsdynamik. Neue physikalische Modellierungskonzepte* (Springer, Berlin, 1997).
8. Henderson, L. F. The statistics of crowd fluids. *Nature* **229**, 381–383 (1971).
9. Henderson, L. F. Sexual differences in human crowd motion. *Nature* **240**, 353–355 (1972).
10. Helbing, D. & Molnár, P. Social force model for pedestrian dynamics. *Phys. Rev. E* **51**, 4282–4286 (1995).
11. Haken, H. *Advanced Synergetics* 2nd edn (Springer, Berlin, 1987).
12. Nicolis, G. & Prigogine, I. *Self-Organization in Nonequilibrium Systems. From Dissipative Structures to Order through Fluctuations* (Wiley, New York, 1977).
13. Schenk, M. *Untersuchungen zum Fußgängerverhalten* Thesis, Univ. Stuttgart (1995).
14. Kayser, D. R., Aberle, L. K., Pochy, R. D. & Lam, L. Active walker models: tracks and landscapes. *Physica A* **191**, 17–24 (1992).
15. Lam, L. Active walker models for complex systems. *Chaos Solitons Fractals* **6**, 267–285 (1995).
16. Schweitzer, F. & Schimansky-Geier, L. Clustering of "active" walkers in a two-component system. *Physica A* **206**, 359–379 (1994).

17. Schimansky-Geier, L., Mieth, M., Rosé, H. & Malchow, H. Structure formation by active Brownian particles. *Phys. Lett. A* **207**, 140–146 (1995).
18. Ben-Jacob, E. *et al.* Generic modelling of cooperative growth patterns in bacterial colonies. *Nature* **368**, 46–49 (1994).
19. Stevens, A. & Schweitzer, F. in *Mechanisms of Cell and Tissue Motion* (eds Alt, W., Deutsch, A. & Dunn, G.) 183–192 (Birkhäuser, Basel, 1997).
20. Eigen, M. & Schuster, P. *The Hypercycle* (Springer, Berlin, 1979).
21. Feistel, R. & Ebeling, W. *Evolution of Complex Systems. Self-Organization, Entropy and Development* (Kluwer, Dordrecht, 1989).
22. Timmermans, H., van der Hagen, X. & Borgers, A. Transportation systems, retail environments and pedestrian trip chaining behaviour: Modelling issues and applications. *Transportation Res. B* **26**, 45–59 (1992).
23. Hillier, B. *Space is a Machine: A Configurational Theory of Architecture* (Cambridge Univ. Press, 1996).

Acknowledgements. We thank F. Schweitzer for discussions, and W. Weidlich and M. Treiber for reviews of the manuscript.

Correspondence should be addressed to D.H. (e-mail: helbing@theo2.physik.uni-stuttgart.de).

Current switching of resistive states in magnetoresistive manganites

A. Asamitsu*, Y. Tomioka*, H. Kuwahara* & Y. Tokura*†

* Joint Research Center for Atom Technology (JRCAT), Tsukuba 305, Japan

† Department of Applied Physics, University of Tokyo, Tokyo 113, Japan

Magnetoresistive devices (based on, for example, magnetic multilayers¹) exhibit large changes in electrical resistance in response to a magnetic field, which has led to dramatic improvements in the data density and reading speed of magnetic recording systems. Manganese oxides having a perovskite structure (the so-called manganites) can exhibit a magnetoresistive response that is many orders of magnitude larger than that found for other materials, and there is therefore hope that these compounds might similarly be exploited for recording applications^{2–11}. Here we show that the switching of resistive states in the manganites can be achieved not only by a magnetic field, but also by an electric field. For manganites of the form $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$, we find that an electrical current (and by implication a static electric field) triggers the collapse of the low-temperature, electrically insulating charge-ordered state to a metallic ferromagnetic state. We suggest that such a phenomenon could be exploited to pattern conducting ferromagnetic domains within an insulating antiferromagnetic matrix, and so provide a route for fabricating micrometre- or nanometre-scale electromagnets.

The carrier-doped manganites with perovskite structure, $\text{R}_{1-x}\text{A}_x\text{MnO}_3$ (R and A being rare-earth and alkaline-earth ions, respectively), are well known as conducting ferromagnets below a Curie temperature T_C (ref. 12). The ferromagnetic interaction, called double-exchange interaction^{13–15}, between localized $3d t_{2g}$ spins (spin quantum number $S = 3/2$) of Mn ions is mediated by conduction $3d e_g$ carriers. Thus, most of the perovskite manganites show a ferromagnetic (FM) ground state when the holes are optimally doped by the chemical substitution of the perovskite A-site. Among various manganites, $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$ is unique, showing insulating behaviour over the whole composition (x) range due to its narrow bandwidth of e_g -electrons⁹. The ground state of $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$ with $x = 0.3$ – 0.5 is a charge-ordered (CO) antiferromagnetic insulator^{16,17}: the real-space ordering of $1:1 \text{ Mn}^{3+}/\text{Mn}^{4+}$ ions occurs at $T_{CO} \approx 230 \text{ K}$ and then local spin moments are antiferromagnetically ordered below a Néel temperature, $T_N \approx 170 \text{ K}$, whereas the ferromagnetic double-exchange interaction between Mn ions is quenched by such a charge-ordering effect. It has been shown^{8,9}, however, that this ‘charge crystal’ can be easily melted into ‘charge liquid’, namely a FM state, by applying a magnetic field, accompanying an antiferromagnetic-to-ferromagnetic (metamagnetic) phase transition as well as a drastic decrease of

the resistivity by more than ten orders of magnitude. Considering the nature of this concomitant metamagnetic and insulator-to-metal transition, we may expect that forcedly moved carriers can revive ferromagnetic interaction and trigger the phase transition without a magnetic field. The presence of large hysteresis or metastable states in the course of the melting transition of the charge-ordered state^{8,9} would also favour such a phase switching.

Crystals of $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$ ($x = 0.3$) were melt-grown by a floating-zone method as described in detail elsewhere^{8,9}. The results of powder X-ray diffraction and electron-probe microanalysis indicated that the crystal is single phase with an orthorhombically distorted ($Pbnm$) lattice. Figure 1 shows the temperature dependence of resistance, $R(T)$, of $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$ ($x = 0.3$) in zero field at various applied voltages. The charge-ordered state of the $x = 0.3$ crystal occurs at $T_{CO} = 230 \text{ K}$, in a similar superlattice pattern to the case of the $x = 0.5$ with the arrangement of $\text{Mn}^{3+}:\text{Mn}^{4+} = 1:1$ (ref. 17), but collapses at a relatively small magnetic field as compared with the $x = 0.5$ crystal⁸. At higher bias voltages, the measured resistance R tends to decrease remarkably as the temperature T is reduced below a spin-canted antiferromagnetic ordering temperature, $T_{CA} = 110 \text{ K}$. The noisy behaviour observed in the $R(T)$ curve below 40 K at a bias voltage of 300 V indicates temporal fluctuations between high-resistivity and low-resistivity states. A bias voltage of 700 V , R at 5 K is as low as $6 \times 10^4 \Omega$, reduced by more than five orders of magnitude from the low-voltage R -value.

As an example, the change of R at 20 K in zero field is shown in

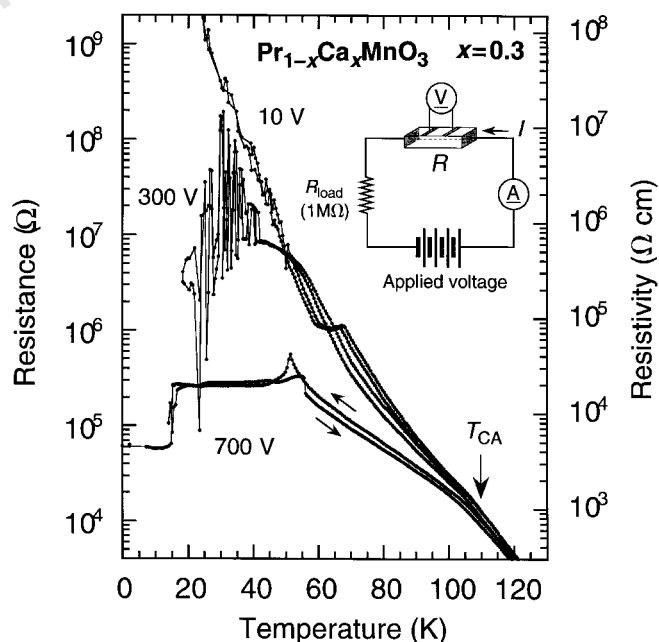


Figure 1 Temperature dependence of resistance of a $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$ ($x = 0.3$) crystal measured at various applied voltages. The measurements were performed using the circuit shown in the inset. A load resistor of $1 \text{ M}\Omega$ was inserted in series to protect the crystal against a burst of current. The crystal was rectangular; 4 mm long, 1 mm wide and 0.6 mm thick. An electrode (for current injection) was made at each end of the crystal (using silver paint) so that the current flow was as uniform as possible. The voltage drop across the crystal was monitored using two other electrodes attached to the surface, which were spaced $\sim 0.9 \text{ mm}$. The resistance was thus measured using the conventional four-wire configuration. The resistivity value on the right ordinate is calculated from current density divided by electric field applied to the crystal, and is meaningful only for the 10 V (ohmic) case. The spin-canted antiferromagnetic ordering temperature, T_{CA} , is 110 K , below which highly nonlinear electric conduction is observed.

Fig. 2 as a function of applied voltage. R is $\sim 10^{10} \Omega$ in the low-voltage region and slowly decreases with increase of bias voltage. Then at a threshold voltage, $V_{\text{threshold}}$, R drops abruptly by over three orders of magnitude. In the decreasing-voltage scan, R shows hysteretic behaviour and then returns to the original value before the switching. We have confirmed that the switching phenomena are not triggered by simple thermal effects or by heating of the crystal. We measured the temperature of the crystal during the voltage scan by use of a carbon resistor that was directly attached to the crystal with grease and was in good thermal contact with the crystal. In the high-resistivity state, the temperature remained exactly that prescribed (20 K) up to the switching voltage; it jumped to 25 K on the switching and in the low-resistivity state tended to increase with applied voltage due to Joule heating. However, the change of the resistance that is expected from the increase of the temperature alone in the ohmic case is less than that observed by about three orders of magnitude (estimated from the resistance data at a bias voltage of 10 V in Fig. 1). Thus simple Joule heating cannot account for the observed switching, which is also evidenced by the following results of the effect of magnetic field on the current switching.

The current-voltage (I - V) characteristics at 20 K in various magnetic fields are shown in Fig. 3. One can see a current jump over three or four orders of magnitude at $V_{\text{threshold}}$. We note that nonlinearity in the I - V curves is observed even before the switching. The experimental conditions of Fig. 3 are indicated by filled

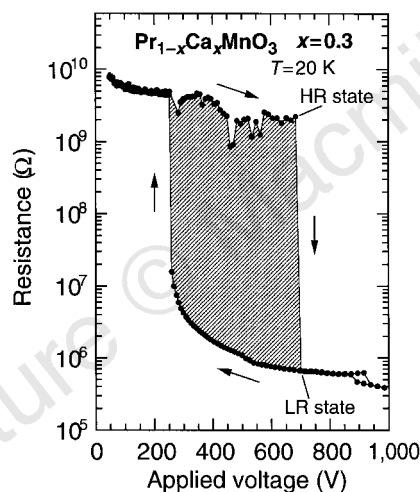


Figure 2 Change of resistance of $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$ ($x = 0.3$) at 20 K as a function of applied voltage, measured using the circuit shown in Fig. 1 inset. In the high-resistivity (HR) state the crystal has a much higher resistance than the load resistor, and the applied voltage is largely across the crystal; in the low-resistivity (LR) state, the voltage is mainly across the load resistor.

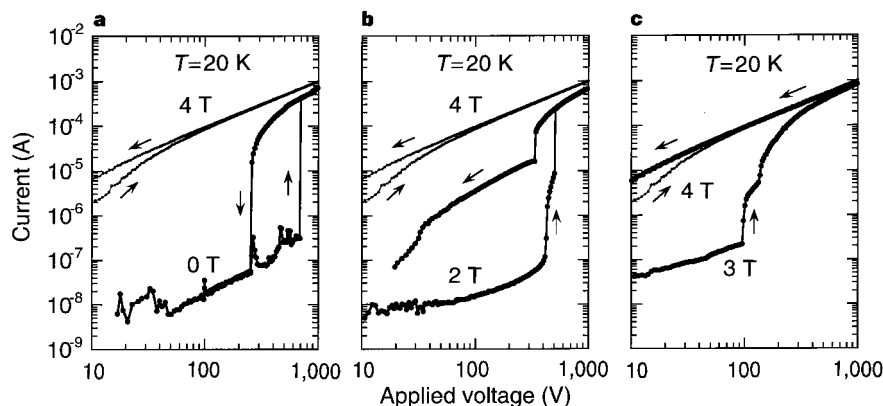


Figure 3 Current-voltage (I - V) characteristics at 20 K in magnetic fields of 0 T (a), 2 T (b) and 3 T (c). The curve at 4 T is shown in each case as the reference ferromagnetic metallic state. The measurements were performed in voltage-scan mode up to 1,000 V using the circuit shown in Fig. 1 inset.

triangles in the magnetoelectronic phase diagram shown in the lower panel of Fig. 4 (refs 8, 9). The field-induced transition between charge-ordered insulator and FM metal is first-order, and so in the hysteretic region (hatched area in Fig. 4) the crystal is bistable for the charge-ordered insulator and the FM metal, depending upon the experimental history on the phase diagram. One can see that the metastable charge-ordered state is most unstable at around $T \approx 20$ K where the magnetic field required to induce a transition between charge-ordered insulator and an FM metal takes a minimum value (~ 4 T). It should be noted, however, that the transition of this metastable state can never be induced by change of temperature at a fixed field, as long as the magnetic field is below ~ 4 T. The threshold voltage, $V_{\text{threshold}}$, as observed in Fig. 3 is plotted in the upper panel of Fig. 4 as a function of magnetic field. Corresponding to the electronic phase diagram in the lower panel of Fig. 4, $V_{\text{threshold}}$ decreases with increase of an applied magnetic field and becomes zero at ~ 4 T.

The I - V curves (Fig. 3) show hysteretic behaviour, with a closed loop when the measurement is carried out at a point far away from the electronic phase boundary (for example, Fig. 3a), meaning that the low-resistivity conduction paths shrink back. Approaching the phase boundary with increase of magnetic field (for example, Fig. 3b,c), the crystal shows an irreversible I - V curve, indicating that the low-resistivity conduction paths are maintained even at zero bias voltage. In Fig. 3c, the I - V curve after the switching to the low-resistivity state (as well as during the decreasing-voltage scan) agrees with the curve for the FM state at 4 T. At zero or low magnetic fields (≤ 2 T), on the other hand, the low-resistivity states after the current switching are not always the bulk metallic states as seen in the I - V curves in the low-resistivity state shown in Fig. 3a,b, and hence the current flow is probably (multi)filamentary. Once the low-resistivity filamentary paths are opened, the current should be centred on these paths. As the charge-ordered insulator/ferromagnetic metal transition is first-order, a critical size of nucleus should be required for the phase transition to expand over the entire sample. When the I - V curve shows hysteretic behaviour with a closed loop, we consider that each path is not extensive enough to induce the phase transition over the entire sample. In the case of the irreversible I - V curves, on the other hand, the phase change to the low-resistivity state is likely to occur as a bulk phenomenon (compare the resistance in the low-resistivity state with the FM-state reference at 4 T; Fig. 3c). The field dependence of $V_{\text{threshold}}$ (Fig. 4) shows clearly that the switching takes place at a smaller bias voltage with increasing magnetic field. This indicates that critical nucleation occurs more easily in the charge-ordered state that is destabilized by a magnetic field.

The current-switching phenomena reported here may involve dielectric breakdown of the charge-ordered state in which, as for Mott insulators more generally, the charge carriers are initially localized at each atomic site owing to electron-electron repulsive interaction. We expect that dielectric breakdown of such states

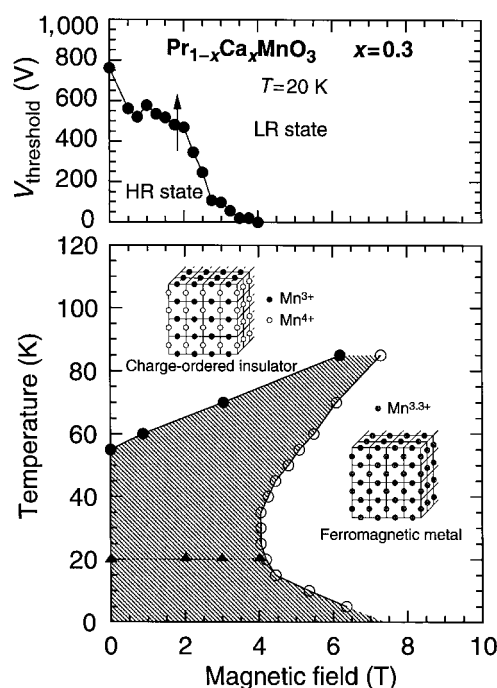


Figure 4 Top panel, threshold voltage at 20 K, $V_{\text{threshold}}$, from a high-resistivity (HR) to a low-resistivity (LR) state, plotted as a function of magnetic field (see Fig. 3). Bottom panel, magneto-electronic phase diagram of $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$ ($x = 0.3$) as a function of magnetic field. Two phase boundaries were determined by field-scan at a fixed temperature. Open circles, critical fields from the charge-ordered insulating (COI) to ferromagnetic metallic (FM) state; filled circles, critical fields for the reverse transition. In the COI state, Mn^{3+} and Mn^{4+} species are regularly ordered on the crystal lattice, while the average valence of Mn ion is 3.3+ in the FM state, as shown in the schematics. The hatched region represents hysteresis where both the COI and FM states are realized, depending on the experimental history on the phase diagram. Experimental conditions of Fig. 3 are indicated by filled triangles.

should immediately lead to the appearance of a metallic state accompanied by a large number of mobile charge carriers (or by a large Fermi surface), and hence is intrinsically different from conventional semiconductors or band insulators. The current-switching phenomenon may be used for the fabrication of electromagnets on a micrometre or nanometre scale, as the switching between the low-resistivity and high-resistivity states in the manganites with charge-ordered instabilities is expected to be accompanied by a metamagnetic transition. This may open the way to a variety of applications such as STM-tip-assisted nanofabrication of ferromagnetic domains in antiferromagnetic matrices. □

Received 27 November 1996; accepted 21 May 1997.

1. Baibich, M. N. *et al.* Giant magnetoresistance of (001)Fe/(001)Cr magnetic superlattices. *Phys. Rev. Lett.* **61**, 2472–2475 (1988).
2. Kusters, R. M., Singleton, D. A., Mcgrevy, R. & Hayes, W. Magnetoresistance measurements on the magnetic semiconductor $\text{Nd}_{0.5}\text{Pb}_{0.5}\text{MnO}_3$. *Physica B* **155**, 362–365 (1989).
3. von Helmont, R., Wecker, J., Holzapfel, B., Schultz, M. & Samwer, K. Giant negative magnetoresistance in perovskite $\text{La}_{0.7}\text{Ba}_{0.3}\text{MnO}_3$ ferromagnetic films. *Phys. Rev. Lett.* **71**, 2331–2333 (1993).
4. Jin, S. *et al.* Thousandfold change in resistivity in magnetoresistive La-Ca-Mn-O films. *Science* **264**, 413–415 (1994).
5. Tokura, Y. *et al.* Giant magnetotransport phenomena in filling-controlled Kondo lattice system: $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$. *J. Phys. Soc. Jpn* **63**, 3931–3935 (1994).
6. Urushibara, A. *et al.* Insulator-metal transition and giant magnetoresistance in $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$. *Phys. Rev. B* **51**, 11103–11109 (1995).
7. Tomioka, Y., Asamitsu, A., Moritomo, Y., Kuwahara, H. & Tokura, Y. Collapse of a charge-ordered state under a magnetic field in $\text{Pr}_{1/2}\text{Sr}_{1/2}\text{MnO}_3$. *Phys. Rev. Lett.* **74**, 5108–5111 (1995).
8. Tomioka, Y., Asamitsu, A., Moritomo, Y. & Tokura, Y. Anomalous magnetotransport properties of $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$. *J. Phys. Soc. Jpn* **64**, 3626–3630 (1995).
9. Tomioka, Y., Asamitsu, A., Kuwahara, H., Moritomo, Y. & Tokura, Y. Magnetic-field-induced metal-insulator phenomena in $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$ with controlled charge-ordering instability. *Phys. Rev. B* **53**, R1689–R1692 (1995).

10. Kuwahara, H., Tomioka, Y., Asamitsu, A., Moritomo, Y. & Tokura, Y. A first-order phase transition induced by a magnetic field. *Science* **270**, 961–963 (1995).
11. Kuwahara, H. *et al.* Striction-coupled magnetoresistance in perovskite-type manganese oxides. *Science* **272**, 80–82 (1996).
12. Jonker, G. H. & Van Santen, J. H. Ferromagnetic compounds of manganese with perovskite structure. *Physica* **16**, 337–349 (1950).
13. Zener, C. Interaction between the d -shells in the transition metals. II. Ferromagnetic compounds of manganese with perovskite structure. *Phys. Rev.* **82**, 403–405 (1951).
14. Anderson, P. W. & Hasegawa, H. Considerations on double exchange. *Phys. Rev.* **100**, 675–681 (1955).
15. de Gennes, P.-G. Effects of double exchange in magnetic crystals. *Phys. Rev.* **118**, 141–154 (1960).
16. Jirak, Z., Krupicka, S., Simsa, Z., Dlouha, M. & Vratislav, Z. Neutron diffraction study of $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$ perovskite. *J. Magn. Magn. Mater.* **53**, 153–166 (1985).
17. Yoshizawa, H., Kawano, H., Tomioka, Y. & Tokura, Y. Neutron-diffraction study of the magnetic-field-induced metal-insulator transition in $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$. *Phys. Rev. B* **52**, R13145–R13148 (1995).

Acknowledgements. This work was performed at JRCAT under the joint research agreement between the NAIR and the ATP, and was supported by NEDO and by a Grant-In-Aid for Scientific Research from the Ministry of Education, Science, and Culture, Japan.

Correspondence and requests for materials should be addressed to A.A. (e-mail: asamitsu@jrcat.or.jp).

Controlled production of aligned-nanotube bundles

M. Terrones*, N. Grobert*, J. Olivares*, J. P. Zhang†, H. Terrones‡, K. Kordatos*, W. K. Hsu*, J. P. Hare*, P. D. Townsend*, K. Prassides*, A. K. Cheetham†, H. W. Kroto* & D. R. M. Walton*

* School of Chemistry, Physics and Environmental Science, University of Sussex, Brighton BN1 9QJ, UK

† Materials Research Laboratory, University of California, Santa Barbara, California 93106, USA

‡ Instituto de Física, UNAM, Apartado Postal 20-364, 01000 México DF, México

Carbon nanotubes^{1,2} might be usefully employed in nanometre-scale engineering and electronics. Electrical conductivity measurements on the bulk material^{3,4}, on individual multi-walled^{5,6} and single-walled⁷ nanotubes and on bundles of single-walled nanotubes^{8,9} have revealed that they may behave as metallic, insulating or semiconducting nanowires, depending on the method of production—which controls the degree of graphitization, the helicity and the diameter. Measurements of Young's modulus show¹⁰ that single nanotubes are stiffer than commercial carbon fibres. Methods commonly used to generate nanotubes—carbon-arc discharge techniques^{1,2,4}, catalytic pyrolysis of hydrocarbons^{11,12} and condensed-phase electrolysis^{13,14}—generally suffer from the drawbacks that polyhedral particles are also formed and that the dimensions of the nanotubes are highly variable. Here we describe a method for generating aligned carbon nanotubes by pyrolysis of 2-amino-4,6-dichloro-*s*-triazine over thin films of a cobalt catalyst patterned on a silica substrate by laser etching. The use of a patterned catalyst apparently encourages the formation of aligned nanotubes. The method offers control over length (up to about 50 μm) and fairly uniform diameters (30–50 nm), as well as producing nanotubes in high yield, uncontaminated by polyhedral particles.

In a typical experiment, a thin film of cobalt (~10–100 nm) was deposited on a silica plate (1 mm thick, 5 mm wide and 20 mm long) using the following technique. A laser beam (Nd:YAG, wavelength 266 nm, 40 mJ per pulse, 10 Hz) was focused (spot size 2 mm outer diameter) on a rotating cobalt target (1.5 cm², 0.25 mm thick) for 10–30 min under vacuum ($\leq 1 \times 10^{-6}$ torr). The silica plate, which was heated to 350 °C during the laser ablation process, was aligned 40 mm from and parallel to the target. Following ablation, the plate was exposed to air and etched with a single laser pulse (5 mJ) using cylindrical lenses (65 mm focal length) to create linear tracks (widths 1–20 μm , lengths ≤ 5 mm; Fig. 1)¹⁶.

A sample of vacuum-dried 2-amino-4,6-dichloro-*s*-triazine (0.02–0.15 g), prepared from cyanuric chloride and ammonia¹⁵, was introduced into one end of a silica tube (6 mm outside

diameter, length 60 cm), and the silica plate coated with etched cobalt was placed face downwards at the other end. The tube was inserted into a two-stage furnace¹⁷ fitted with independent temperature controllers (Fig. 2). Argon gas ($20 \text{ cm}^3 \text{ min}^{-1}$) was passed through the system and the temperature of the second furnace was set at 950°C . The temperature of the first furnace was then raised at $15^\circ\text{C min}^{-1}$ to 200°C and then, by 100°C increments (5 min per increment) to $1,000^\circ\text{C}$. After 5 min at this temperature, the first furnace was allowed to cool to room temperature, but the second furnace was maintained at 950°C for an additional 15 min in order to complete the annealing process. At this stage, the silica plate was covered with dark tracks visible to the naked eye; those areas where the cobalt film had not been laser-etched were transparent (see below).

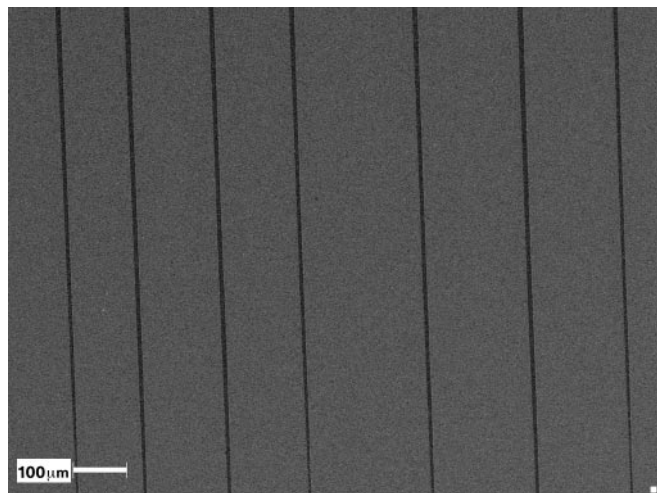


Figure 1 SEM image showing uniform tracks etched by the laser beam. It is believed that along these channels/tracks uniform cobalt nanoparticles are deposited evenly by the laser striking over the thin film. Scale bar, $100 \mu\text{m}$.

A series of silica plates, treated in this way, were coated with gold and examined by scanning electron microscopy, SEM (Leo 5420 operated at 20 keV). The black deposit on a second series of plates was removed by scraping, dispersed in acetone, and analysed by the following techniques: transmission electron microscopy (TEM, using a JEOL JEM 100CX at 100 keV), high resolution TEM (HRTEM, using a JEOL JEM 2010 at 200 keV and a JEM 4000 at 400 keV, and a Gatan GIF system operated at 200 kV with 0.3-eV dispersion for fine structure study) and energy dispersive X-ray analysis, EDX (using a NORAN Instruments detector attached to the latter microscope). Residues from the acetone exit bubblers (Fig. 2) were analysed by mass spectrometry, MS (VG autospec, electron impact, 70 eV).

SEM studies of the gold-coated plates revealed the presence of

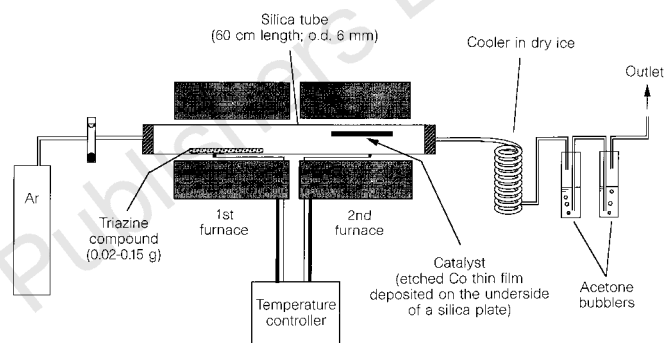
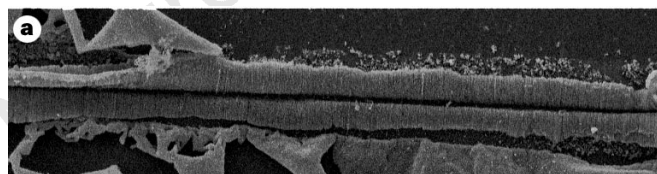
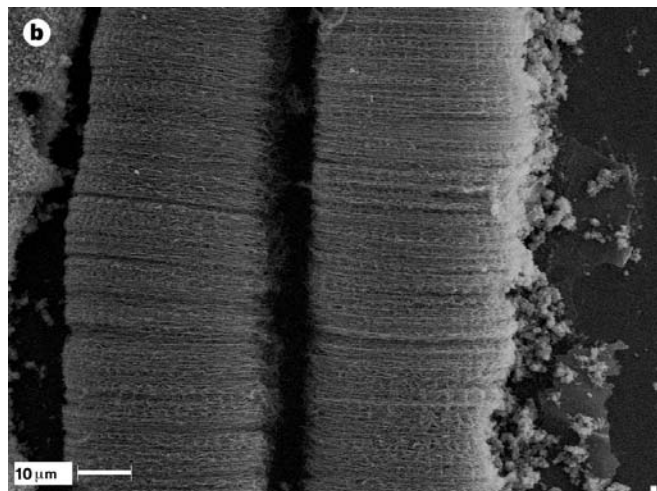


Figure 2 Pyrolysis device, illustrating the two-stage furnace system.

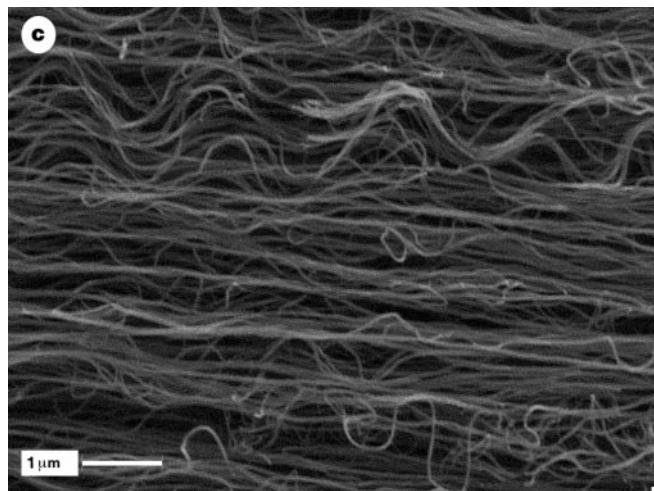


100 μm

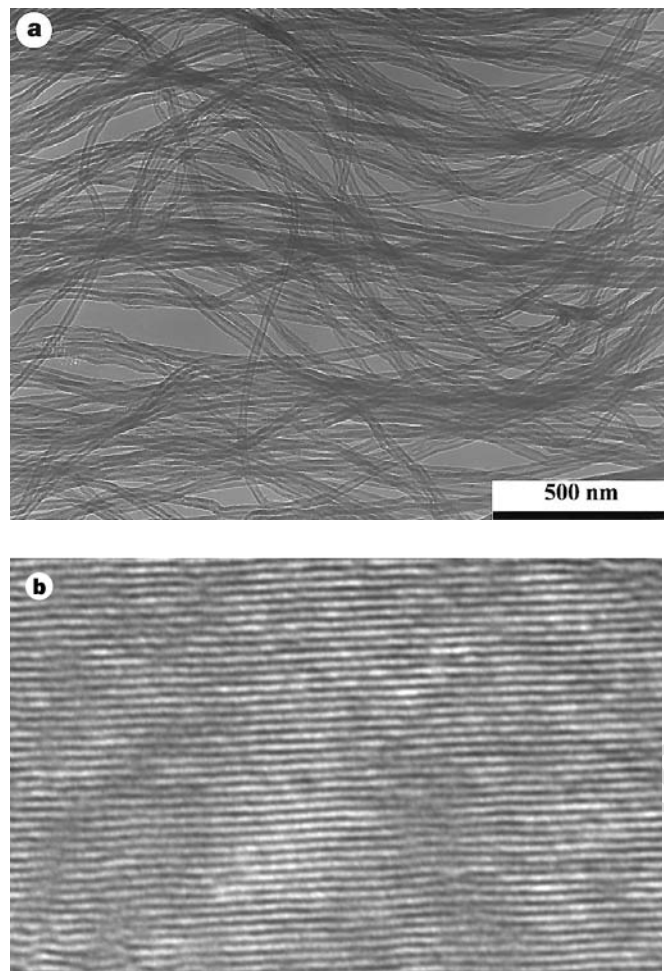


10 μm

Figure 3 SEM images of aligned nanotube bundles. **a**, Low-magnification of adjacent bundles in which nanotubes appear aligned. Scale bar, $100 \mu\text{m}$. **b**, **c**, Higher magnification of one bundle, showing aligned nanotubes of uniform length ($40 \mu\text{m}$) and diameter (30–50 nm). Scale bars: in **b**, $10 \mu\text{m}$; in **c**, $1 \mu\text{m}$.



1 μm



nanotube bundles closely aligned with the nanotracks (length 1–5 mm) created by laser etching. The tubes within these bundles were of uniform length ($\leq 50 \mu\text{m}$) and external diameter ($\sim 30\text{--}50 \text{ nm}$; Fig. 3). The residual cobalt on the plate was removed, probably by the action of HCl, and was collected in the exit bubblers as cobalt chloride. We note that no traces of encapsulated or polyhedral particles were detected. Aligned-nanotube films also were observed in other etched regions.

TEM and HRTEM observations (Fig. 4a, b) confirm the presence of multi-layered graphitic tubules (30–50 nm outer diameter, 60–80 layers). In most cases, cobalt (particles $\leq 50 \text{ nm}$ diameter) was detected by EDX analysis within the nanotube tips. These particles appear to be responsible for the nanotube growth, but they were absent from a significant number (5%) of the closed end-caps. Occasionally, substantial sections of tubes were filled with cobalt. Analyses (using electron energy-loss spectroscopy, EELS) show that the nanotubes consist of pure carbon accompanied by traces of nitrogen ($< 2\text{--}5\%$). Sharp ionization edges at 284 and 291 eV correspond to π^* and σ^* features associated with sp^2 hybridized carbon. A very broad weak feature at 400 eV may be due to nitrogen, generated during triazine decomposition and trapped inside the tubules. Both EELS and EDX measurements indicate that chlorine is entirely absent.

The deposition of a cobalt thin film on silica by laser ablation seems to be an efficient method for producing a uniform distribution of the metal catalyst. Subsequent laser-etching generates tracks or areas free of cobalt (Fig. 1), leaving cobalt particles evenly positioned along the edges of the eroded tracks or stripes¹⁶. These particles ($\leq 50 \text{ nm}$), which exhibit a large surface/volume ratio,

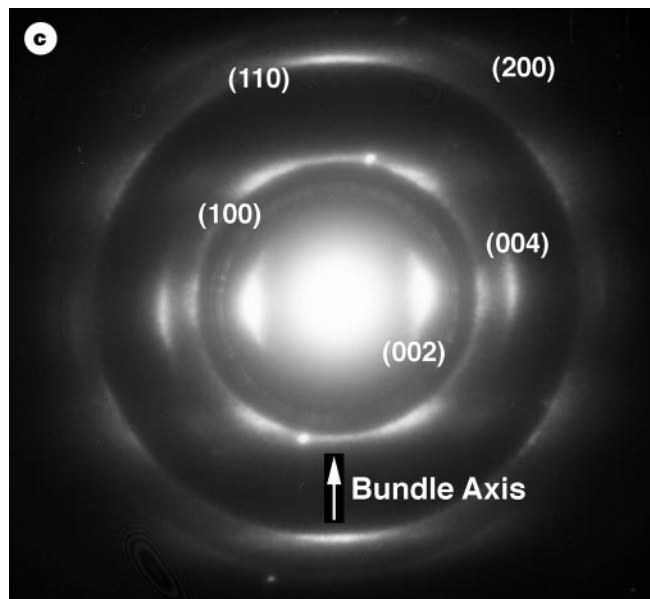


Figure 4 a, TEM image of a typical region filled with pure nanotubes dispersed by sonication. Encapsulated and polyhedral particles or other graphitic nanostructures are absent. Scale bar, 500 nm. **b**, HRTEM image showing good graphitization within the nanotubes (interlayer spacing $\sim 3.4 \text{ \AA}$). **c**, Electron diffraction pattern from the group of nanotubes, exhibiting a collective feature. The tubes are grown in the direction indicated by the headed arrow. The centre of the first ring corresponds to $3.4 \text{ \AA}$ in real space, the rings revealing the presence of hexagonal graphite (armchair arrangement) which has an orientational relationship with the nanotubes. The walls of the tubes are parallel to (0001) reflections.

appear to be responsible for the carbon agglomeration and the unique nanotube growth behaviour reported here. An important factor in our experiments is the ratio of organic precursor (2-amino-4,6-dichloro-*s*-triazine) to cobalt. If this exceeds 2,700 : 1 (by weight) the cobalt is completely removed as cobalt chloride and the catalyst is destroyed.

It is not clear at this stage why the particle alignment occurs. The nanotubes appear to grow preferentially through the aligned cobalt crystals. Overcrowding may be responsible for simultaneous tube growth from the surface. The experiment was conducted with the cobalt catalyst on the lower (inverted) silica surface. No aligned tube growth occurred when the cobalt-coated/etched surface was on the upper face, so that gravitational effects may well be significant.

An electron diffraction pattern (Fig. 4c) recorded for a group of nanotubes reveals a highly ordered graphitic arrangement within the bundles, especially with respect to the (001) plane. In addition, the outer rings indicate the presence of hexagonal graphite, commensurate with a preferential direction for nanotube growth (related to a non-helical arrangement corresponding to an armchair configuration). These patterns are usually observed (20–30%) within analysed samples.

While this work was in progress a report appeared describing the large-scale synthesis of aligned carbon nanotubes by passage of acetylene over iron nanoparticles embedded in mesoporous silica¹⁸. Pyrolytic formation of nanotubes in high yield that are substantially free from pyrolytic carbon overcoatings have been reported by Hyperion Inc.^{19,20}, but full details of the methods and product characterization have not been published, and the nanotubes do not appear to be aligned²¹. □

Received 13 March; accepted 12 May 1997.

1. Iijima, S. Helical microtubules of graphitic carbon. *Nature* **354**, 56–58 (1991).
2. Ebbesen, T. W. & Ajayan, P. M. Large scale synthesis of carbon nanotubes. *Nature* **358**, 220–222 (1992).
3. De Heer, W. A., Chatelain, A. & Ugarte, D. A carbon nanotube field-emission electron source. *Science* **270**, 1179–1180 (1995).
4. Terrones, M. *et al.* Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials Vol. 2 (eds Kadish, K. M. & Ruoff, R. S.) 599–620 (Electrochem. Soc., Pennington, NJ, 1995).
5. Dai, H. J., Wong, E. W. & Lieber, C. M. Probing electrical transport in nanomaterials: conductivity of individual carbon nanotubes. *Science* **272**, 523–526 (1996).
6. Ebbesen, T. W. *et al.* Electrical conductivity of individual carbon nanotubes. *Nature* **382**, 54–56 (1996).
7. Tans, S. J. *et al.* Individual single-wall carbon nanotubes as quantum wires. *Nature* **386**, 474–477 (1997).
8. Bockrath, M. *et al.* Single-electron transport in ropes of carbon nanotubes. *Science* **275**, 1922–1925 (1997).
9. Thess, A. *et al.* Crystalline ropes of metallic carbon nanotubes. *Science* **273**, 483–487 (1996).
10. Treacy, M. M. J., Ebbesen, T. W. & Gibson, J. M. Exceptionally high Young's modulus observed for individual nanotubes. *Nature* **381**, 678–680 (1996).
11. Amelinckx, S. *et al.* A formation mechanism for catalytically grown helix-shaped graphite nanotubes. *Science* **265**, 635–639 (1994).
12. Endo, M. *et al.* Pyrolytic carbon nanotubes from vapor-grown carbon fibres. *Carbon* **33**, 873–881 (1995).
13. Hsu, W. K. *et al.* Condensed-phase nanotubes. *Nature* **377**, 687 (1995).
14. Hsu, W. K. *et al.* Electrolytic formation of carbon nanostructures. *Chem. Phys. Lett.* **261**, 161–166 (1996).
15. Thurston, J. T. *et al.* Cyanuric chloride derivatives I. Aminochloro-s-triazines. *J. Am. Chem. Soc.* **73**, 2981–2983 (1951).
16. Chrisey, D. B. & Hubler, G. K. (eds) *Pulsed Laser Deposition of Thin Films* (Wiley, New York, 1994).
17. Terrones, M. *et al.* Pyrolytically grown B₂C₃N₂ nanomaterials: nanofibres and nanotubes. *Chem. Phys. Lett.* **257**, 576–582 (1996).
18. Li, W. Z. *et al.* Large synthesis of aligned carbon nanotubes. *Science* **274**, 1701–1703 (1996).
19. Tennent, H. G., Barber, J. J. & Hoch, R. US Patent No. 5578543 (1996).
20. Hausslein, R. W. Commercial manufacture and uses of carbon nanotubes. *187th Mtg of the Electrochem. Soc. (Abstr.)* 175 (Electrochem. Soc., Pennington, NJ, 1995).
21. Niu, C., Sichel, E. K., Hoch, R., Moy, D. & Tennent, H. High power electrochemical capacitors based on carbon nanotube electrodes. *Appl. Phys. Lett.* **70**, 1480–1482 (1997).

Acknowledgements. We thank J. Thorpe, D. P. Randall, S. Tehuacanero, R. Hernández, P. Mexia, R. Guadiana and L. Rendón for providing electron microscope facilities, and D. Bernaerts for discussions. We thank CONACYT-México (M.T. and H.T.), the ORS scheme for scholarships (M.T. and W.K.H.), DGAPA-UNAM IN 107-296 (H.T.), EU-TMR grant (J.O.), the Royal Society (London) and the EPSRC for financial support.

Correspondence and requests for materials should be addressed to D.R.M.W. (e-mail: d.walton@sussex.ac.uk).

Chemical properties of element 106 (seaborgium)

M. Schädel^{*}, W. Brüche^{*}, R. Dressler[†], B. Eichler[†], H. W. Gäggeler[‡], R. Günther[§], K. E. Gregorich^{||}, D. C. Hoffman[¶], S. Hübener[#], D. T. Jost[†], J. V. Kratz[§], W. Paulus[§], D. Schumann[☆], S. Timokhin^{}, N. Trautmann[§], A. Türler[†], G. Wirth^{*} & A. Yakushev^{**}**

^{*} Gesellschaft für Schwerionenforschung, 64291 Darmstadt, Germany

[†] Labor für Radio- und Umweltchemie, Paul Scherrer Institut, 5232 Villigen, Switzerland

[‡] Department für Chemie und Biochemie, Universität Bern, 3012 Bern, Switzerland

[§] Institute für Kernchemie, Universität Mainz, 55099 Mainz, Germany

^{||} Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

[¶] Glenn T. Seaborg Institute for Transactinium Science, Livermore, California 94551, USA

[#] Institut für Radiochemie, Forschungszentrum Rossendorf, 01314 Dresden, Germany

[☆] Institut für Analytische Chemie, Technische Universität Dresden, 01062 Dresden, Germany

^{**} Flerov Laboratory of Nuclear Reactions, Joint Institute of Nuclear Research, Dubna, Russia

The synthesis, via nuclear fusion reactions, of elements heavier than the actinides, allows one to probe the limits of the periodic table as a means of classifying the elements. In particular, deviations in the periodicity of chemical properties for the heaviest elements are predicted as a consequence of increasingly strong relativistic effects on the electronic shell structure^{1–7}. The trans-

actinide elements have now been extended up to element 112 (ref. 8), but the chemical properties have been investigated only for the first two of the transactinide elements, 104 and 105 (refs 9–19). Those studies showed that relativistic effect render these two elements chemically different from their lighter homologues in the same columns of the periodic table (Fig. 1). Here we report the chemical separation of element 106 (seaborgium, Sg) and investigations of its chemical behaviour in the gas phase and in aqueous solution. The methods that we use are able to probe the reactivity of individual atoms, and based on the detection of just seven atoms of seaborgium we find that it exhibits properties characteristic of the group 6 homologues molybdenum and tungsten. Thus seaborgium appears to restore the trends of the periodic table disrupted by relativistic effects in elements 104 and 105.

Calculations of the electron configurations of heavy atoms^{1–5} have predicted that sudden changes in the structure of the electron shells may appear due to strongly increasing relativistic effects. These relativistic effects are proportional to the square of the nuclear charge, which attracts electrons in spherically symmetric orbitals (*s* and *p*_{1/2}) most strongly to the nucleus. This, in turn, means that the nuclear charge is more efficiently screened, thus allowing expansion of the non-spherical *d* and *f* orbitals. Because the chemical behaviour of an element is strongly dependent on the electronic configuration, such relativistic effects can lead to unexpected chemical properties^{6,7}.

Studies of the chemical properties of elements 104 (rutherfordium, Rf) and 105 (hahnium, Ha; the name dubnium has also been proposed, but has yet to be approved by the International Union of Pure and Applied Chemistry) were full of surprises^{18,19}. The non-tantalum-like behaviour of hahnium in aqueous solution^{10,11,20}, for example, and its similarity to niobium and/or protactinium, depending on its chemical environment, demonstrated that the chemical properties cannot be reliably extrapolated from the trends observed in its lighter homologues. Such surprises have also been seen in thermo-chromatographic^{9,21} and gas-chromatographic experiments^{12,13,15,16}. Therefore, it is of interest to investigate whether the chemical properties of element 106 (seaborgium, Sg) resemble those of the lighter homologues in group 6 (molybdenum and tungsten) or those of the pseudo-group-6 element uranium.

We synthesized the most neutron-rich seaborgium isotopes^{22,23}, ²⁶⁵Sg and ²⁶⁶Sg, in a nuclear fusion reaction between ²²Ne ions from the GSI UNILAC accelerator and a ²⁴⁸Cm target with a rate of the order of one atom per hour (ref. 24). The ²⁶⁵Sg and ²⁶⁶Sg nuclei were knocked out of the target foil and were stopped in helium gas loaded with tiny (0.1–1 µm) solid particles (aerosols). Within about three seconds, the helium transported the reaction products—attached to aerosols—along capillary tubes to two different sets of chemical devices. To provide conclusive evidence that a seaborgium atom had passed through the chemical separation procedures, the experiments were designed to detect the characteristic α-decay chains of the isotope ²⁶⁵Sg, and the corresponding daughter nuclides ²⁶¹Rf and ²⁵⁷No (ref. 24). In addition, fission fragments were measured from spontaneous fission decay, which would arise from ²⁶²Rf, the α-decay product of ²⁶⁶Sg. An earlier attempt to perform a chemical separation of element 106 fell short of unambiguously showing that the observed, by itself unspecific, spontaneous fission decay originated from an isotope of element 106 (ref. 25). We have applied two chemical separation techniques, one probing the formation of volatile oxychlorides in a gas-chromatographic experiment, and another one probing the formation of oxo- or oxyfluoride complexes in aqueous solution by liquid chromatography.

In classical gas chromatography, a substance under investigation is introduced into a flowing stream, and the time taken for the sample to emerge from chromatographic system is measured. In our experiment, in contrast, the nuclear reaction products were continuously supplied and separated in OLGA III—the on-line gas chemistry apparatus¹⁷. This technique uses the half-life of the

Figure 1 shows the periodic table of elements. The actinides and lanthanides are shown separately below the main table. Element 106, seaborgium (Sg), is highlighted in group 6, period 7.

Figure 1 Periodic table of the elements. The arrangement of the actinides reflects the fact that the first actinide elements still resemble, to a decreasing extent, the chemistry of the other groups: Th the fourth group below Hf, Pa the fifth group below Ta, and U the sixth group below W. The known transactinide elements 104 to 112 take the positions from below Hf in group 4 to below Hg in group 12. Element 106, seaborgium (Sg), the heaviest element chemically investigated, is placed in group 6.

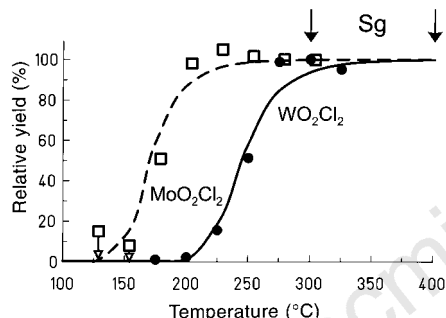


Figure 2 Measured relative yields of the compounds MoO_2Cl_2 (open squares) and WO_2Cl_2 (filled circles) versus temperature of the gas-chromatographic column. The lines are the result of Monte Carlo simulations which model the behaviour of the compounds during separation. Arrows indicate the temperatures at which seaborgium was separated.

nuclide under investigation as an 'internal' clock for the chromatographic process. Just half of the atoms introduced into the gas-chromatographic column will emerge from the other end when the time taken for an atom to pass through the column (retention time) corresponds to the nuclide's half-life. In this way, yields (that is, the ratio of the quantity introduced into the column to the quantity detected after the column) can be converted into retention times when the column is run at various temperatures. Figure 2 shows the behaviour of compounds of the short-lived isotopes of the homologues molybdenum and tungsten under the experimental conditions used for the separation of seaborgium. Reactive gases—chlorine saturated with thionyl chloride, and a little oxygen—were added to the chemically inert carrier-gas helium. Given a sufficiently high temperature, the volatile oxychloride compounds form. Thermodynamic calculations indicate the formation of dioxydichlorides as the most stable compound. Figure 2 shows that MoO_2Cl_2 is more volatile than WO_2Cl_2 because the former passes through the chromatography column at a lower temperature. The experiment with seaborgium atoms was carried out at temperatures of 300 and 400 °C, high enough for detection of the volatile oxychlorides—assuming that these seaborgium compounds do not exhibit unexpectedly low volatilities.

In the course of the OLGA experiment, the chemically separated volatile compounds leaving the end of the quartz gas-chromatography column are deposited on aerosol particles and rapidly

transported in a gas stream to a detector. Here, about six seconds after their formation, the seaborgium compounds are deposited on thin foils, mounted on a revolving wheel. The wheel is rotated in a ten-second cycle so as to position the samples between detector pairs, which detect the α -particles and spontaneous-fission fragments, and measure their characteristic energies and time sequences. With OLGA III, it was possible for the first time to detect unambiguously three α -decay chains of ^{265}Sg after solid/gas-phase chemical separation (Fig. 3). An additional α -decay, which was followed 2.8 seconds later by a spontaneous fission, can be assigned to the decay of the neighbouring isotope ^{266}Sg .

Our chemical results show that, under the given conditions, element 106 forms a volatile oxychloride at 300 and 400 °C. At each temperature, two α -decay chains were observed. Thermodynamic calculations indicate that SgO_2Cl_2 is formed, in analogy to MoO_2Cl_2 and WO_2Cl_2 . This agrees with the expected behaviour from an extrapolation in group 6 of the periodic table and with theoretical calculations⁷.

Further chromatographic separations were carried out with single seaborgium atoms to determine its ion valency and complex formation in aqueous solution²⁴. Using the separation apparatus ARCA²⁶ (automated rapid chemistry apparatus), investigations with molybdenum and tungsten showed that hexavalent ions could be eluted from cation exchange columns with dilute nitric/hydrofluoric acid. This chemical system makes use of characteristic differences in the formation of cationic, neutral and anionic complexes with F^- ions between elements of groups 3, 4, 5 and 6. Di- and trivalent actinides, group 4 elements, and the pseudo-group-6 element uranium, present as UO_2^{2+} , are retained on the cation exchange resin. However, the group 6 elements molybdenum and tungsten form anions of the type MO_4^{2-} , MO_3F^- or MO_2F_3^- (where M indicates a metal ion)²⁷. Moreover, the formation of a neutral compound like MO_2F_2 can not be excluded. Anionic and neutral species readily elute from the column.

Theoretical calculations of the electronic structure and redox potentials predict that, within the series of MO_4^{2-} ions formed in aqueous solution, SgO_4^{2-} will be the most stable²⁸. From this, we expect that in very dilute hydrofluoric acid seaborgium forms SgO_4^{2-} which should follow the lighter homologues molybdenum and tungsten in the chemical separation.

During the course of more than 5,000 liquid-chromatographic separations (elution time, 10 seconds each), it was possible to detect three α -decay chains of the daughter nuclei of ^{265}Sg —namely, the α -decay of ^{261}Rf and ^{257}No —in the seaborgium fraction²⁴ (Fig. 3). They constitute unambiguous proof that seaborgium has passed

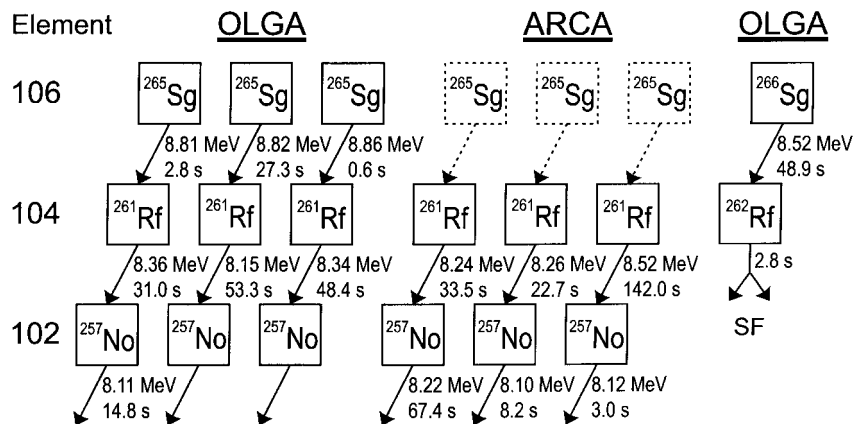


Figure 3 The observed nuclear decay chains from the seaborgium isotopes ^{265}Sg and ^{266}Sg , which allowed an unambiguous identification of seaborgium after chemical separation with ARCA (automated rapid chemistry apparatus) and OLGA (on-line gas chemistry apparatus). The α -decay energies are given in MeV, and the observed life-times in seconds.

through the column, given that both of these nuclides could only be present in the seaborgium fraction due to the α -decay of ^{265}Sg . Isotopes of the elements 104 and 102 formed directly or from decay of seaborgium before chemical separation were chemically separated. With a 90% probability, the parent nuclei had already decayed into the daughter nucleus ^{261}Rf in the time between the end of chemical separation and the commencement of measurement about 28 seconds later. This is very likely in view of the short life-times measured for the ^{265}Sg decays with OLGA (Fig. 3).

The first liquid-chromatographic separation of element 106 shows that at least a substantial fraction of the formed seaborgium behaves similarly to its lighter homologues molybdenum and tungsten, that is, that its behaviour is typical of a hexavalent ion belonging to group 6 of the periodic table. Presumably, seaborgium forms SgO_4^{2-} or a neutral complex.

Both our isothermal gas chromatographic and liquid chromatographic separations clearly indicate that seaborgium behaves similarly to its lighter homologues molybdenum and tungsten, and its behaviour is typical for a group 6 element of the period table. These results support the assumption that the chemistry of elements 107 to 112 will be homologous to that of the group 7 to 12 elements rhenium to mercury—if increasingly strong relativistic effects do not alter the chemical properties to such an extent that they are no longer predictable from such simple extrapolations, as seen for elements 104 and 105. □

Received 27 January; accepted 12 May 1997.

1. Fricke, B. & Greiner, W. On the chemistry of superheavy elements around $Z = 164$. *Phys. Lett. B* **30**, 317–319 (1969).
2. Desclaux, J.-P. & Fricke, B. Relativistic prediction of the ground state of atomic lawrencium. *J. Phys.* **41**, 943–946 (1980).
3. Glebov, V. A., Kasztura, L., Nefedov, V. S. & Zhuikov, B. L. Is element 104 (kurchatovium) a p-element? II. Relativistic calculations of the electronic atomic structure. *Radiochim. Acta* **46**, 117–121 (1989).
4. Johnson, E., Fricke, B., Keller, O. L., Nestor, C. W. & Tucker, T. C. Ionization potentials and radii of atoms and ions of element 104 (unnilquadium) and of hafnium ($2+$) derived from multiconfiguration Dirac-Fock calculations. *J. Chem. Phys.* **93**, 8041–8050 (1990).
5. Eliav, E., Kaldor, U. & Ishikawa, Y. Ground state electron configuration of rutherfordium: role of dynamic correlation. *Phys. Rev. Lett.* **74**, 1079–1082 (1995).
6. Pyykkö, P. Relativistic effects in structural chemistry. *Chem. Rev.* **88**, 563–594 (1988).
7. Pershina, V. G. Electronic structure and properties of the transactinides and their compounds. *Chem. Rev.* **96**, 1977–2101 (1996).
8. Hofmann, S. *et al.* The new element 112. *Z. Phys. A* **354**, 229–230 (1996).
9. Zvara, I., Belov, V. Z., Domanov, V. P. & Shalavskii, M. R. Chemical isolation of nilsbohrium as ekatantalum in the form of the anhydrous bromide. II. Experiments with a spontaneously fissioning isotope of nilsbohrium. *Sov. Radiochem.* **18**, 328–334 (1976).
10. Gregorich, K. E. *et al.* Aqueous chemistry of element 105. *Radiochim. Acta* **43**, 223–231 (1988).
11. Kratz, J. V. *et al.* Chemical properties of element 105 in aqueous solution: halide complex formation and anion exchange into triisooctyl amine. *Radiochim. Acta* **48**, 121–133 (1989).
12. Gäggeler, H. W. *et al.* Gas phase chromatography experiments with bromides of tantalum and element 105. *Radiochim. Acta* **57**, 93–100 (1992).
13. Gäggeler, H. W. On-line gas chemistry experiments with transactinide elements. *J. Radioanal. Nucl. Chem.* **183**, 261–271 (1994).
14. Czerwinski, K. R. *et al.* Solution chemistry of element 104: Part I. Liquid–liquid extractions with triisooctylamine. *Radiochim. Acta* **64**, 23–28 (1994); Solution chemistry of element 104: Part II. Liquid–liquid extractions with tributylphosphate. *Radiochim. Acta* **64**, 29–35 (1994).
15. Türlér, A. *et al.* On-line gas phase chromatography with chlorides of niobium and hahnium (element 105). *Radiochim. Acta* **73**, 55–66 (1996).

16. Kadkhodayan, B. *et al.* On-line gas chromatographic studies of chlorides of rutherfordium and homologs Zr and Hf. *Radiochim. Acta* **72**, 169–178 (1996).
17. Türlér, A. Gas phase chemistry experiments with transactinide elements. *Radiochim. Acta* **72**, 7–17 (1996).
18. Schädel, M. Chemistry of the transactinide elements. *Radiochim. Acta* **70/71**, 207–223 (1996).
19. Hoffman, D. C. Chemistry of the heaviest elements. *Radiochim. Acta* **72**, 1–6 (1996).
20. Zimmermann, H. P. *et al.* Chemical properties of element 105 in aqueous solution: back extraction from triisooctyl amide into 0.5 M HCl. *Radiochim. Acta* **60**, 11–16 (1993).
21. Zvara, I. *et al.* Gas chromatography and thermochromatography in the study of transuranium elements. *Sov. Radiochem.* **16**, 709–715 (1974).
22. Loughheed, R. W. *et al.* in *Proc. Int. Conf. Actinides-93* (eds Clark, D. L., Hobart, D. E. & Fuger, F.) 161 (Elsevier, Amsterdam, 1994); *J. Alloy Compounds* **213/214**, 61–66 (1994).
23. Lazarev, Yu. A. *et al.* Discovery of enhanced nuclear stability near the deformed shells $N = 162$ and $Z = 108$. *Phys. Rev. Lett.* **73**, 624–627 (1994).
24. Schädel, M. *et al.* First aqueous chemistry with seaborgium (element 106). *Radiochim. Acta* (in the press).
25. Timokhin, S. N., Yakushev, A. B., Honggui Xu, Perelygin, V. P. & Zvara, I. Chemical identification of element 106 by thermochromatography. *J. Radioanal. Nucl. Chem. Lett.* **212**, 31–34 (1996).
26. Schädel, M. *et al.* ARCA II—a new apparatus for fast, repetitive HPLC separations. *Radiochim. Acta* **48**, 171–176 (1989).
27. Tytko, K.-H. & Gras, D. in *Gmelin, Handbook of Inorganic Chemistry; Molybdenum Suppl. Vol. B 3b* (Springer, Heidelberg, 1989).
28. Pershina, V. & Fricke, B. The electronic structure of the group 6 oxyanions $[\text{MO}_4]^{2-}$, where $M = \text{Cr}, \text{Mo}, \text{W}$, and element 106. *Radiochim. Acta* **65**, 13–17 (1994).

Acknowledgements. We are indebted to the Division of Chemical Sciences, Office of Basic Energy Research, US Department of Energy, for making the ^{248}Cm target material through the transplutonium element production program at the Oak Ridge National Laboratory. We also thank the staff and crew of the GSI UNILAC, the TRIUMF reactor, the Philips-cyclotron at PSI, and the MPI Heidelberg tandem accelerator for their help. This work was supported in part by the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF), the Swiss National Science Foundation, and the Chemical Sciences Division of the Office of Basic Energy Sciences, US Department of Energy.

Correspondence should be addressed to M.S. (e-mail: m.schaedel@gsi.de).

Vegetation and climate change in northwest America during the past 125 kyr

Cathy Whitlock & Patrick J. Bartlein

Department of Geography, University of Oregon, Eugene, Oregon 97403, USA

Vegetation records spanning the past 21 kyr in western North America display spatial patterns of change that reflect the influence of variations in the large-scale controls of climate¹. Among these controls are millennial-scale variations in the seasonal cycle of insolation and the size of the ice sheet, which affect regional climates directly through changes in temperature and net radiation, and indirectly by shifting atmospheric circulation. Longer vegetation records provide an opportunity to examine the regional response to different combinations of these large-scale controls, and whether non-climatic controls are important. But most of the longer North American records^{2,3} are of insufficient quality to allow a robust test, and the long European records^{4–9} are in regions where the vegetation response to climate is often

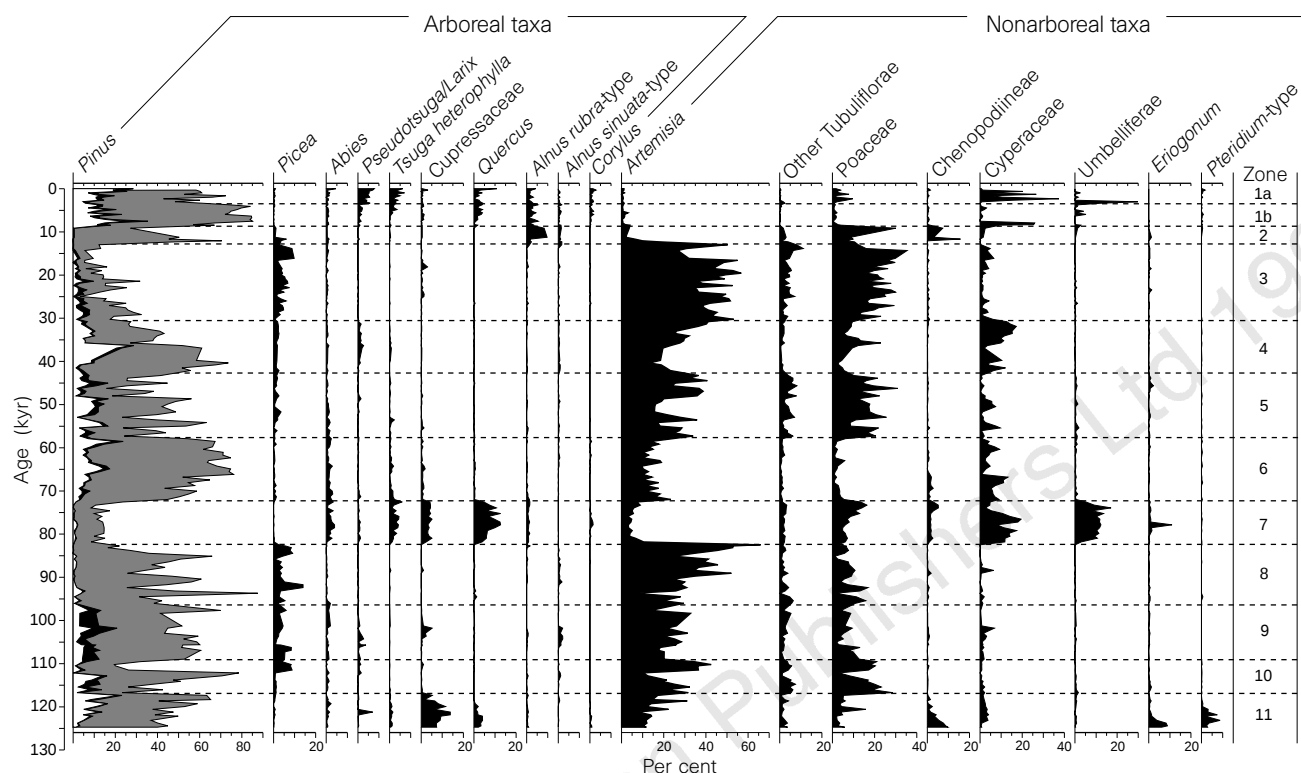


Figure 1 Pollen percentage diagram of selected taxa from Carp Lake cores 90 and 93. *Pinus* includes diploxylon-type pollen (open curve) from *P. contorta* or *P. ponderosa*, haploxylon-type pollen (black curve) from *P. monticola* or *P. albicaulis*, and indeterminate *Pinus* pollen (shaded curve). A minimum of 350 terrestrial pollen grains was counted for each level and used as the pollen sum in calculating percentages.

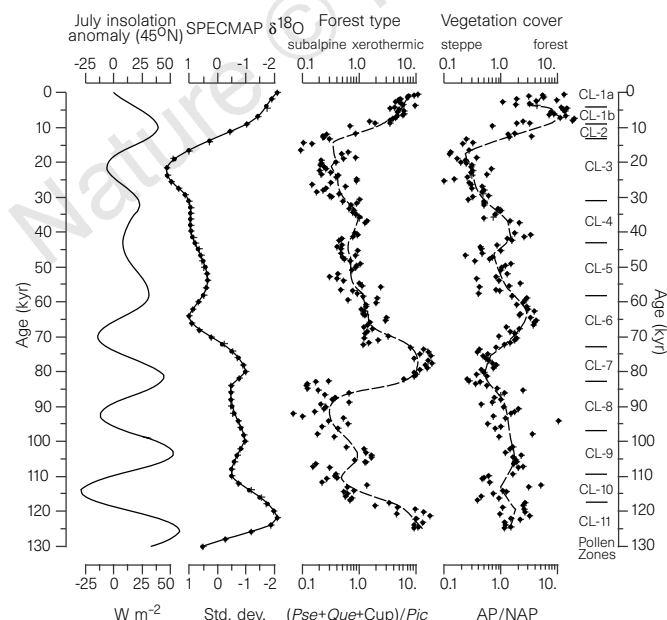


Figure 2 Large-scale controls (July insolation anomalies at latitude 45°N and global ice volume) and Carp Lake vegetation indices. The ratios of the sum of *Pseudotsuga/Larix*, *Quercus* and *Cupressaceae* pollen percentages to *Picea* pollen percentages ($Pse + Que + Cup/Pic$) and of total arboreal to total non-arboreal pollen (AP/NAP) provide an indication of vegetation type and openness. The smoothed curves were drawn using a 'lowess' procedure³⁰.

difficult to separate from the response to ecological and anthropogenic controls. Here we present a 125-kyr record of vegetation and climate change for the forest/steppe border of the eastern Cascade Range, northwest America. Pollen data disclose alternations of forest and steppe that are consistent with variations in summer insolation and global ice-volume, and vegetational transitions correlate well with the marine isotope-stage boundaries. The close relationship between vegetation and climate beyond the Last Glacial Maximum provides evidence that climate variations are the primary cause of regional vegetation change on millennial timescales, and that non-climatic controls are secondary.

Regional vegetation history is generally regarded as a response to climate variations operating on different spatial and temporal scales¹⁰. One approach to understanding this hierarchy is the comparison of model simulations and data syntheses (for example, palaeoclimate model simulations of the past 21 kyr, produced by general circulation models, are compared with palaeoenvironmental maps constructed from palaeoecological data¹¹). In northwest America, such comparisons reveal the influence of the Laurentide ice sheet and the seasonal cycle of insolation on regional climate¹. Model simulations of full-glacial conditions^{12,13} suggest that the ice sheet caused cooling of the northern mid-latitudes, steepened the latitudinal temperature gradient, and shifted the jet stream south of its present position. In addition, glacial anticyclonic circulation led to weakened westerlies, prevailing easterlies, and enhanced cold dry conditions along the southern ice margin. Simulations of the late-glacial period (16–10 kyr ago) suggest that a smaller ice sheet caused cool humid conditions as a result of a latitudinal temperature depression and a northwards shift of the jet. In the early Holocene

Table 1 Ages and tephra information used to establish age–depth relations at Carp Lake

Depth* (m)	Uncalibrated ¹⁴ C age (yr BP)	Calibrated ¹⁴ C age (kyr)†	Lab no., tephra identification‡, or δ ¹⁸ O age equivalent	Core
1.3	0	0	Sediment surface	90
2.15–2.16	3,450 ± 450	3.691	Mt St Helens, Ye	85
3.10–3.20	5,820 ± 50	6.667	QL-1640	85
3.64–3.68	6,800 ± 200	7.581	Mazama ash	85
3.70–3.77	8,760 ± 40	9.754	WIS-1460	85
3.82–3.92	9,470 ± 100	10.472	WIS-1468	85
4.46–4.56	9,730 ± 400	10.961	QL-1641	85
5.61–5.71	16,050 ± 400	18.925	QL-1642	85
6.15–6.25	19,790 ± 190	23.45	Beta-57036	90
6.76–6.86	21,100 ± 400	24.706	QL-1643	85
8.30–8.40	26,200 ± 200	29.59	QL-1646	85
9.10–9.20	32,700 ± 450	35	QL-1603	85
9.25–9.35	32,760 ± 420	35	Beta-57037	90
19.11–19.50	~100,000	100§	Mt St Helens, WA-5C	90
15.15		73.9	Stage 4/5a	93
23.15		125	Stage 5e	93

The following are regression equations:

$$\text{Age (kyr)} = -5.9284 + 4.1052\text{depth} + 0.0095\text{depth}^2 + 0.0032\text{depth}^3 \quad (1)$$

$$\text{Age (kyr)} = -3.3742 + 2.6245\text{depth} + 0.2186\text{depth}^2 - 0.0040\text{depth}^3 \quad (2)$$

Equation (2) incorporates δ¹⁸O ages for the pollen equivalents of stage 4/5a boundary and stage 5e (see ref. 18).

* Radiocarbon dates and tephra dates were used from core 85 (ref. 14) to a core depth of 9.8 m (below water surface), core 90 to a depth of 20.19 m, and core 93 from 20.19 to 23.15 m. Correlation among cores was based on lithology.

† Radiocarbon yr were converted to calendar yr (kyr) by use of the CALIB3 program for the past 20,000 ¹⁴C-yr BP (ref. 27), the U-Th calibration for ages between 20,000 and 30,000 ¹⁴C-yr BP (ref. 28), and the geomagnetic record for the period from 30,000 to 50,000 ¹⁴C-yr BP (ref. 29). Tephra ages are reported in refs 15, 16.

‡ Tephra identifications provided by A. Sarna-Wojcicki, US Geol. Survey.

§ Based on unpublished data (A. Sarna-Wojcicki) and extrapolation from a series of thermoluminescence (TL) dates in a nearby loess section (ref. 16). The age assignment is preliminary until further dating is conducted.

epoch, greater-than-present summer insolation increased summer temperatures, decreased effective precipitation, and indirectly strengthened the eastern Pacific subtropical high-pressure system, which intensified drought.

Comparisons between model results and data for the past 21 kyr encompass only a small subset of the large-scale variations that have occurred in the Quaternary period. Records that preserve evidence of the consequences of combinations of large-scale controls before the Last Glacial Maximum offer an alternative way to examine the hierarchy of controls of past climate variations. In such an approach, the past 21 kyr offer hypotheses about the relative influence of ice-sheet size and insolation variations that can be tested with the record from earlier periods.

A pollen record from Carp Lake (latitude 45° 55' N, longitude 120° 53' W, altitude 714 m above sea level), a crater lake in the eastern Cascade Range, provides a vegetation history of the past 125 kyr. The site lies at the lower altitudinal limit of *Pinus ponderosa* forest near *Artemisia* steppe, which is a sensitive ecotone for recording past variations in temperature and effective precipitation. Cores were collected in 1985¹⁴, 1990 and 1993 to yield a 23.15-m-long record, and pollen was analysed from 196 samples taken at regular intervals. A preliminary age versus depth relationship was established based on radiocarbon dates and the ages of tephra layers^{15,16} in the cores (Table 1, equation (1)). The pollen percentage record was divided into assemblage zones based on a constrained cluster analysis¹⁷ (Fig. 1). Pollen zone boundaries were assigned independently of the marine δ¹⁸O stratigraphy¹⁸, yet their ages matched the isotopic stage boundaries surprisingly well. The original chronology was revised by incorporating in the regression analysis the ages of marine isotope stage boundary 4/5a and the height of stage 5e (Table 1, equation (2)).

Interpretations of past vegetation and climate rested on a comparison of the pollen stratigraphy with modern pollen spectra¹⁹, vegetation²⁰, and climate²⁰ (Table 2). Ratios of *Pseudotsuga* + *Quercus* + Cupressaceae/*Picea* provided information on forest type, with high values indicating low-elevation forest and warm, effectively dry conditions. These taxa were chosen because their pollen distribution today is sharply delimited by elevation, unlike *Pinus*, *Artemisia*, *Alnus* and *Tsuga*, which have widely dispersed pollen. Arboreal/non-arboreal pollen ratios were an index of vege-

tation cover and were used to separate periods of forest from steppe or tundra.

The pollen record is composed of taxa that grow today in the Cascade Range and Columbia basin. The previous interglaciation (CL-11) was characterized by open forests of *Pinus*, *Quercus* and *Juniperus* and by steppe; climate during this time was warmer and drier than most of the Holocene. Alternations of open *Pinus* and *Pinus-Picea* forest and closed mixed-conifer forest occurred from 83 to 117 kyr ago (CL-8 to CL-10) during cooler-than-present conditions. An open forest of *Pseudotsuga* or *Larix*, *Tsuga*, *Abies* (probably *A. grandis*), *Quercus* and Cupressaceae was present between 73 and 83 kyr ago (CL-7). The presence of *Quercus*, Chenopodiaceae, and steppe herbs suggests that summers were warm, and the expansion of *Tsuga*, *Abies* and Poaceae indicates increased summer moisture. Cool conditions during the early and middle Wisconsin interval (CL-6 to CL-4) supported mixed conifer forests of *Pinus*, *Picea* and sometimes *Pseudotsuga* or *Larix*. The last glaciation (CL-3) featured cold dry steppe, and an expansion of *Picea* in the late-glacial. Temperate steppe was present in the early Holocene (CL-2), and Carp Lake dried intermittently¹⁴. The middle Holocene (CL-1b) was characterized by *Pinus* forest, and modern vegetation was established in the past 3,900 yr (CL-1a).

The close agreement between the pollen zonation and the marine δ¹⁸O stratigraphy, which is a proxy for global ice volume, suggests that vegetation changes were controlled by global variations in the climate system. Most of the variability at Carp Lake relates to the slowly changing insolation and ice-volume record, leaving little to be accounted for by higher-frequency variations, such as Heinrich events²¹. The synchronicity of the changes also suggests negligible lags between large-scale climatic forcing and local vegetational response. Apparently, species kept pace with millennial-scale climate changes by expanding their ranges short distances from discontinuous refugia.

Shifts in forest types parallel variations in both the marine δ¹⁸O record²² and anomalies in July insolation²³ for the past 125 kyr (Fig. 2). Periods of low ice volume and summer insolation maxima featured xerothermic taxa. Periods of moderate ice volume and summer insolation minima supported subalpine conifers. When the combination of large-scale controls was unique, so too was the vegetation. For example, anomalous conditions during the full-glacial

Table 2 Carp Lake vegetation and climate history

Pollen zone age and depth	Oxygen isotope stage and age (ref. 18)	Vegetation	Inferred climate
CL-1a 0–3.9 kyr (1.30–2.35 m)	1 0–14.1 kyr	Development of the modern forest in the Late Holocene. <i>Pinus ponderosa</i> and <i>Pseudotsuga menziesii</i> became the dominant trees, with <i>Alnus</i> , <i>Corylus</i> and other shrubs. Mesic locations supported <i>Abies grandis</i> , <i>Pinus monticola</i> and <i>Tsuga heterophylla</i> . <i>Quercus</i> woodland was established at lower treeline. Poaceae pollen indicates the establishment of grassland at lower elevations and in forest openings.	Modern
CL-1b 3.9–9.1 kyr (2.35–3.70 m)		Middle Holocene forest of <i>Pinus ponderosa</i> and <i>Quercus</i> woodland.	Warmer drier than present
CL-2 9.1–13.2 kyr (3.70–4.65 m)		Early Holocene steppe vegetation with Poaceae and Chenopodiaceae. <i>Alnus</i> was present in riparian settings.	Warmer drier than present
CL-3 13.2–30.9 kyr (4.65–8.25 m)	2 14.1–27.6 kyr	Cold steppe with <i>Artemisia</i> , Poaceae and herbs during the full-glacial period. Temperate aquatic taxa (for example, <i>Nuphar</i> , <i>Sparganium</i> -type, <i>Myriophyllum</i> , <i>Polygonum amphibium</i> -type and <i>Typha</i>) are absent. <i>Polygonum bistortoides</i> -type pollen (not shown) indicates subalpine or alpine conditions. Populations of <i>Picea engelmannii</i> near the site, especially in late-glacial period.	Coldest driest period
CL-4 30.9–43.1 kyr (8.25–10.4 m)	3 27.6–58.9 kyr	Open forest with mixture of low- and high-elevation species. <i>Picea</i> and <i>Abies</i> present in mesic settings, and lower or drier forests supported <i>Pinus contorta</i> or <i>P. ponderosa</i> , <i>Pseudotsuga</i> or <i>Larix occidentalis</i> , and <i>Abies grandis</i> . <i>Picea</i> and <i>Artemisia</i> pollen are more abundant at the bottom of the zone.	Cooler drier than present
CL-5 43.1–58.0 kyr (10.4–12.85 m)		Open forest or forest-steppe vegetation, dominated by <i>Pinus ponderosa</i> and/or <i>P. contorta</i> with lesser amounts of <i>Picea</i> and <i>Abies</i> . Pollen of Poaceae, <i>Artemisia</i> and other herbs suggests forest openings or nearby steppe.	Cooler than present, relatively humid
CL-6 58.0–72.7 kyr (12.85–15.15 m)	4 58.9–73.9 kyr	Closed <i>Pinus</i> forest with some <i>Picea</i> and <i>Abies</i> . <i>Artemisia</i> percentages indicate dry forest openings.	Cooler drier than present
CL-7 72.7–82.8 kyr (15.15–16.70 m)	5a 73.9–85.1 kyr	Open forest of <i>Pseudotsuga/Larix</i> , <i>Tsuga heterophylla</i> , <i>Abies</i> , <i>Quercus</i> and Cupressaceae; the first four taxa are typical of low elevations in the western Cascade Range. Cupressaceae pollen may have come from <i>Thuja plicata</i> or <i>Chamaecyparis nootkatensis</i> , both mesophytes, or <i>Juniperus occidentalis</i> , a xerophyte. Chenopodiaceae, Poaceae, <i>Pteridium</i> , Umbelliferae and <i>Eriogonum</i> were present.	Warmer wetter summers than present
CL-8 82.8–96.8 kyr (16.70–18.83 m)	5b 85.1–93.6 kyr	Open forest of <i>Picea engelmannii</i> , <i>Pinus contorta</i> and possibly <i>P. ponderosa</i> . <i>Alnus sinuata</i> grew in areas of seepage. <i>Artemisia</i> percentages imply nearby steppe or forest openings.	Cooler than present, dry
CL-9 96.8–109.5 kyr (18.83–20.75 m)	5c 93.6–107 kyr	Closed forest with <i>Pinus</i> , <i>Pseudotsuga/Larix</i> , <i>Abies</i> and <i>Picea</i> , similar to montane forests in the Cascades and northeastern Oregon. <i>Pinus</i> pollen was probably from <i>P. contorta</i> , <i>P. ponderosa</i> , and <i>P. monticola</i> or <i>P. albicaulis</i> .	Warmer summers than present, cool humid winters
CL-10 109.5–117.3 kyr (20.75–21.95 m)	5d 107–116.7 kyr	Open <i>Pinus</i> forest with <i>Picea</i> , <i>Abies</i> , <i>Artemisia</i> and Poaceae.	Cooler than present, humid
CL-11 117.3–124.9 kyr (21.95–23.15 m)	5e 116.7–133 kyr	Open forest during previous interglaciation. <i>Pinus</i> is from <i>P. ponderosa</i> or <i>P. contorta</i> . <i>Quercus</i> suggests dry woodland at lower elevations. Cupressaceae is ascribed to <i>Juniperus</i> . Chenopodiaceae, <i>Eriogonum</i> , <i>Artemisia</i> and Poaceae values imply steppe element. <i>Brasenia schreberi</i> (not shown) occurs in this zone and the Holocene only.	Warmer drier than Holocene

led to no-analogue steppe (CL-3). During the last half of the previous interglaciation, summer insolation exceeded that of the early Holocene and global ice volume was at its lowest; these conditions fostered a warm, dry climate and an expansion of xerophytic forest (CL-11). The interglacial period had greater forest cover than the early Holocene, and the record does not show the strong climate variability during stage 5e noted in the Greenland ice-core record²⁴ and some European pollen records²⁵. Stage 5a featured an unusual combination of high summer insolation, modest global ice volumes, and cooler-than-present sea surface temperatures in the north Pacific Ocean²⁶. High summer insolation probably increased growing-season temperature, but a cool ocean and a steepened temperature gradient may have lessened drought and allowed the eastward shift of mesophytic taxa at the expense of *Pinus* and *Artemisia* (CL-7).

Differences in vegetational succession among interglacial and interstadial periods have received considerable attention, particularly in Europe where several long pollen sequences have been described^{14–9}. The differences are often attributed to singular circumstances during particular periods, including non-climatic factors such as refugial location, dispersal-rate differences among taxa, and biotic interactions. Like European records, Carp Lake shows the

individualistic response of vegetation during the Late Quaternary, but the close correspondence of the vegetational variations and the large-scale controls of climate indicates that non-climatic variables were of little or no importance on millennial timescales. Carp Lake offers compelling evidence that interglacial and interstadial vegetation assemblages owe their distinctiveness primarily to the effects of global variations in climate. □

Received 23 December 1996; accepted 13 May 1997.

- Thompson, R. E., Whitlock, C., Bartlein, P. J., Harrison, S. & Spaulding, W. G. in *Global Climates since the Last Glacial Maximum* (eds Wright, H. E. Jr et al.) 468–513 (Univ. Minnesota, Minneapolis, 1993).
- Heusser, L. E. & King, J. E. in *Vegetation History* (eds Huntley, B. & Webb, T.) 193–236 (Kluwer, Dordrecht, 1988).
- Adam, D. P. Correlations of the Clear Lake, California, core CL-73-7 pollen sequence with other climate records. *Geol. Soc. Am. Spec. Pap.* **214**, 81–96 (1988).
- Watts, W. A. in *Vegetation History* (eds Huntley, B. & Webb, T.) 155–192 (Kluwer, Dordrecht, 1988).
- Watts, W. A., Allen, J. R. M. & Huntley, B. Vegetation history and paleoclimate of the last glacial-interglacial period at Lago Grande di Monticchio, southern Italy. *Quat. Sci. Rev.* **15**, 113–154 (1996).
- Pons, A., Guiot, J., de Beaulieu, J. L. & Reille, M. Recent contributions to the climatology of the last glacial interglacial cycle based on French pollen sequences. *Quat. Sci. Rev.* **11**, 439–448 (1992).
- Guiot, J., Pons, A., de Beaulieu, J. L. & Reille, M. A 140,000-year continental climate reconstruction from two European pollen records. *Nature* **338**, 309–313 (1989).
- Tzekadis, P. C., Bennett, K. D. & Magri, D. Climate and the pollen record. *Nature* **370**, 513 (1994).
- Tzekadis, P. C. & Bennett, K. D. Interglacial vegetation succession: a view from southern Europe. *Quat. Sci. Rev.* **14**, 967–982 (1995).
- Huntley, B. & Webb, T. *Vegetation History* (Kluwer, Dordrecht, 1988).
- Wright, H. E. Jr et al. *Global Climates since the Last Glacial Maximum* (Univ. Minnesota, Minneapolis, 1993).

12. Kutzbach, J. E. & Guetter, P. J. The influence of changing orbital patterns and surface boundary conditions on climate simulations for the past 18,000 years. *J. Atmos. Sci.* **43**, 1726–1759 (1986).
13. Hall, N. M. J., Valdes, P. J. & Dong, B. The maintenance of the last great ice sheets: a UGAMP GCM study. *J. Clim.* **9**, 1004–1009 (1996).
14. Barnosky, C. W. A record of late-Quaternary vegetation from the southwestern Columbia Basin, Washington. *Quat. Res.* **23**, 109–122 (1985).
15. Sarna-Wojcicki, A. J. in *Late Quaternary Environments of the United States* Vol. 2 (ed. Wright, H. E. Jr) 52–77 (Univ. Minnesota, Minneapolis, 1983).
16. Berger, G. W. & Busacca, A. J. Thermoluminescence dating of Late Pleistocene loess and tephra from eastern Washington and southern Oregon and implications for the eruptive history of Mount St. Helens. *J. Geophys. Res.* **100**, 22361–22374 (1995).
17. Grimm, E. C. in *Vegetation History* (eds Huntley, B. & Webb, T.) 53–76 (Kluwer, Dordrecht, 1988).
18. Martinson, D. G. *et al.* Age dating and the orbital theory of the ice ages: development of a high resolution 0 to 300,000-year chronostratigraphy. *Quat. Res.* **27**, 1–29 (1987).
19. Mack, R. N. & Bryant, V. M. Jr Modern pollen spectra from the Columbia Basin, Washington. *Northwest Sci.* **48**, 183–194 (1974).
20. Franklin, J. F. & Dyrness, C. T. *Natural Vegetation of Oregon and Washington* (Oregon State Univ., Corvallis, 1988).
21. Bond, G. *et al.* Correlations between climate records from North Atlantic sediments and Greenland ice. *Nature* **365**, 143–147 (1993).
22. Imbrie, J. *et al.* in *Milankovitch and Climate* (eds Berger, A., Imbrie, J., Hays, J., Kukla, G. & Saltzman, B.) 269–305 (Reidel, Dordrecht, 1984).
23. Berger, A. & Loutre, M. F. Insolation values for the last 10 million years. *Quat. Sci. Rev.* **10**, 297–317 (1991).
24. Greenland Ice-core Project (GRIP) Members. Climate instability during the last interglacial period recorded in the GRIP ice core. *Nature* **364**, 203–207 (1993).
25. Thouveny, N. *et al.* Climate variations in Europe over the past 140 kyr deduced from rock magnetism. *Nature* **371**, 503–506 (1994).
26. Morley, J. J., Pisias, N. G. & Leinen, M. Late Pleistocene time series of atmospheric and oceanic variables recorded in sediments from the subarctic Pacific. *Paleoceanography* **2**, 49–62 (1987).
27. Stuiver, M. & Reimer, P. J. Extended ¹⁴C data base and revised CALIB 3.0 ¹⁴C age calibration program. *Radiocarbon* **35**, 215–230 (1993).
28. Bard, E., Hamelin, B., Fairbanks, R. G. & Zindler, A. Calibration of the ¹⁴C timescale over the past 30,000 years using mass spectrometric U–Th ages from Barbados corals. *Nature* **345**, 405–410 (1990).
29. Mazaud, A., Laj, C., Bard, E., Arnold, M. & Tric, A. E. Geomagnetic field control of ¹⁴C production over the last 80 kyr: implications for the radiocarbon time-scale. *Geophys. Res. Lett.* **18**, 1885–1888 (1991).
30. Cleveland, W. S. *Visualizing Data* (Hobart, Summit, 1993).

Acknowledgements. We thank A. Sarna-Wojcicki for tephra identifications, R. J. Nickmann for help with the pollen analysis, and J. Guiot and R. E. Gresswell for reviews. The work was supported by the NSF and the Westinghouse-Hanford Paleoclimate Program.

Correspondence and requests for materials should be addressed to C.W. (e-mail: whitlock@oregon.uoregon.edu).

Endemic African mammals shake the phylogenetic tree

Mark S. Springer*, Gregory C. Cleven*, Ole Madsen†, Wilfried W. de Jong†‡, Victor G. Waddell§, Heather M. Amrine* & Michael J. Stanhope§

* Department of Biology, University of California, Riverside, California 92521, USA

† Department of Biochemistry, University of Nijmegen, PO Box 9101, 6500 HB Nijmegen, The Netherlands

‡ Institute for Systematics and Population Biology, University of Amsterdam, PO Box 94766, 1090GT Amsterdam, The Netherlands

§ Biology and Biochemistry, Queen's University, 97 Lisburn Road, Belfast BT9 07BL, UK

The order Insectivora, including living taxa (lipotyphlans) and archaic fossil forms, is central to the question of higher-level relationships among placental mammals¹. Beginning with Huxley², it has been argued that insectivores retain many primitive features and are closer to the ancestral stock of mammals than are other living groups³. Nevertheless, cladistic analysis suggests that living insectivores, at least, are united by derived anatomical features⁴. Here we analyse DNA sequences from three mitochondrial genes and two nuclear genes to examine relationships of insectivores to other mammals. The representative insectivores are not monophyletic in any of our analyses. Rather, golden moles are included in a clade that contains hyraxes, manatees, elephants, elephant shrews and aardvarks. Members of this group are of presumed African origin^{5,6}. This implies that there was an extensive African radiation from a single common ancestor that gave rise to ecologically divergent adaptive types. 12S ribosomal RNA transversions suggest that the base of this radiation occurred during Africa's window of isolation in the Cretaceous period

before land connections were developed with Europe in the early Cenozoic era.

Relationships among orders of placental mammals have proved difficult to resolve¹. To extend the available mitochondrial (mt) sequences, a 2.6-kilobase (kb) segment containing the 12S rRNA, valine transfer RNA, and 16S rRNA genes was sequenced for nine taxa to generate a data set that is representative of 12 of the 18 placental orders and all three insectivore suborders⁴. Phylogenetic analyses provide strong support for well-established mammalian clades such as carnivores, hominoids, and Cetacea plus Artiodactyla (Fig. 1a). In agreement with other molecular studies^{7–10} that included an assortment of taxa, most interordinal associations are not resolved at bootstrap values >75%. However, the mtDNA data do provide strong support for the association of the two paenungulates (hyrax, manatee) together, and of these with elephant shrews, aardvarks and golden moles (Fig. 1a and Table 1). The association of hyraxes with proboscideans and sirenians was suggested by Cope¹¹. A competing hypothesis is an association of hyraxes with perissodactyls¹². Our results agree with earlier protein^{13,14} and DNA studies^{7–10} supporting Cope's paenungulate hypothesis. In addition to bootstrap support, T-PTP¹⁵ and Kishino–Hasegawa¹⁶ tests also support paenungulate monophyly (Table 2). Anatomical data provide some evidence that aardvarks and/or elephant shrews may be related to paenungulates^{17,18} but suggest other hypotheses as well: for example, six osteological features are putative synapomorphies uniting elephant shrews with lagomorphs and rodents¹⁹. All the available sequence data, including amino-acid sequences^{13,14}, DNA sequences for three nuclear genes^{8–10}, and the present mitochondrial genes, support an association of aardvarks and elephant shrews with paenungulates. What is most unexpected is that golden moles, a family of insectivores, are also part of this clade. 12S rRNA sequences earlier suggested an association of golden moles with paenungulates, but did not provide convincing bootstrap support for this hypothesis⁷. Our expanded data set demonstrates that insectivores are not monophyletic (Table 2)

Table 1 Bootstrap support for select clades based on different methods

	Clade	
	Paenungulata	Paenungulata + aardvark + elephant shrew + golden mole
Mitochondrial DNA		
Parsimony	99	95
Transversion parsimony	64	90
Minimum evolution		
Tamura–Nei I	100	92
Tamura–Nei II	100	78
Logdet	99	90
Maximum likelihood	100	100
vWF		
Parsimony		
All positions	49	99
1st and 2nd positions	24	65
3rd positions	51	93
Transversion parsimony	30	95
Minimum evolution		
Tamura–Nei I	37	99
Tamura–Nei II	30	99
Logdet	43	97
Maximum likelihood	78	100
A2AB		
Parsimony		
All sites	71	88
1st and 2nd positions	49	81
3rd positions	31	67
Transversion parsimony	71	54
Minimum evolution		
Tamura–Nei I	83	84
Tamura–Nei II	28	25
Logdet	79	78
Maximum likelihood	81	89

Only two of the three paenungulate orders were represented among the mitochondrial and A2AB sequences. Tamura–Nei²⁷ I and II distances were calculated by using an equal-rates assumption and a gamma-distribution of rates, respectively.

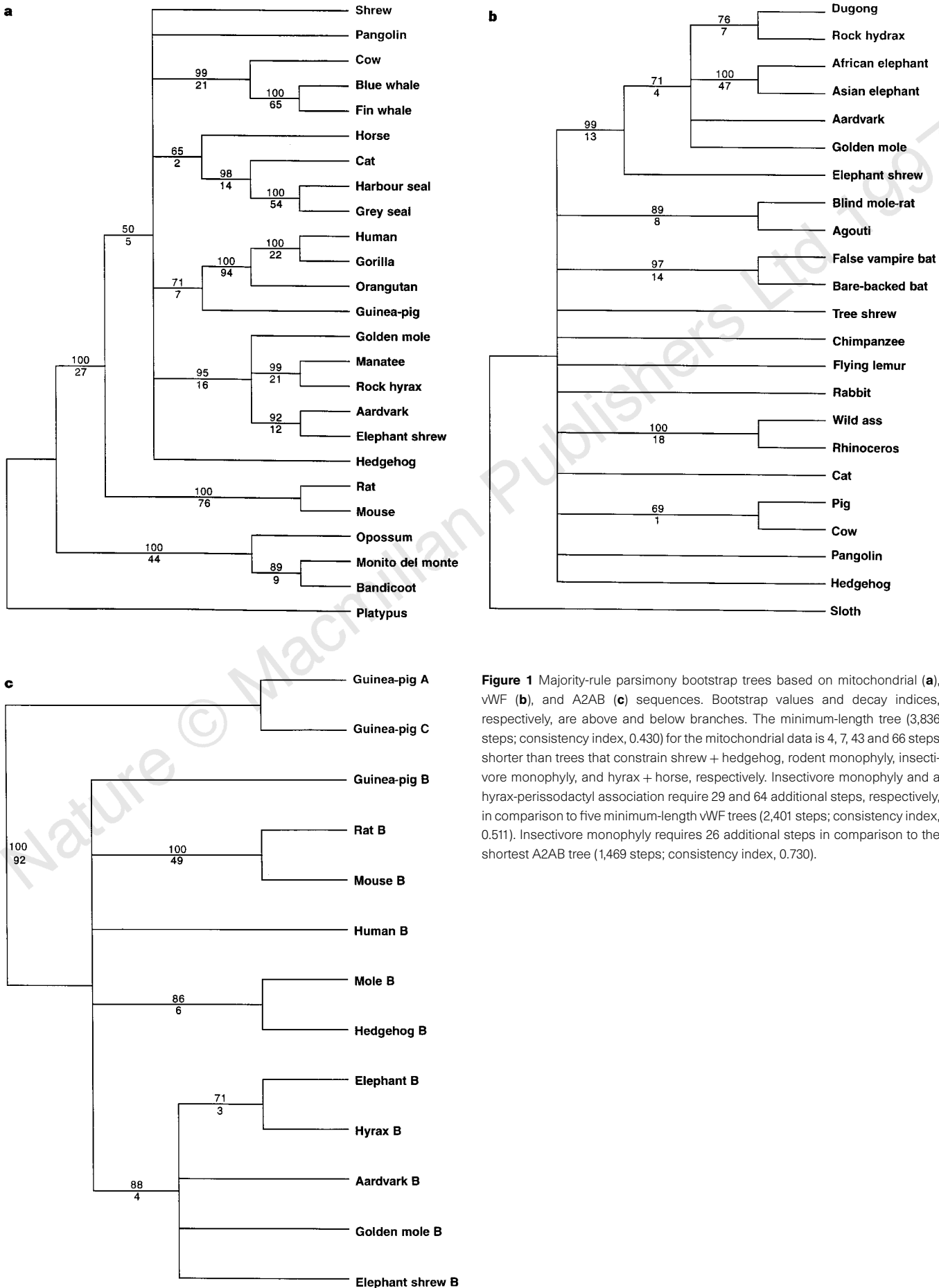


Figure 1 Majority-rule parsimony bootstrap trees based on mitochondrial (**a**), vWF (**b**), and A2AB (**c**) sequences. Bootstrap values and decay indices, respectively, are above and below branches. The minimum-length tree (3,836 steps; consistency index, 0.430) for the mitochondrial data is 4, 7, 43 and 66 steps shorter than trees that constrain shrew + hedgehog, rodent monophyly, insectivore monophyly, and hyrax + horse, respectively. Insectivore monophyly and a hyrax-perissodactyl association require 29 and 64 additional steps, respectively, in comparison to five minimum-length vWF trees (2,401 steps; consistency index, 0.511). Insectivore monophyly requires 26 additional steps in comparison to the shortest A2AB tree (1,469 steps; consistency index, 0.730).

Table 2 Significance levels of T-PTP and Kishino–Hasegawa tests

Constraint	Mitochondrial DNA			vWF			A2AB		
	T-PTP	KH-P	KH-L	T-PTP	KH-P	KH-L	T-PTP	KH-P	KH-L
Perissodactyls + hyracooids	0.01	0.0011–0.0022	<0.0001	0.00	<0.0001	<0.0001	MT	MT	MT
Insectivore monophyly	0.05	<0.0001	0.0001	0.01	0.0311	0.0477	0.00	0.0002–0.0067	0.0001

In each case, trees with constraints were compared against either minimum length (T-PTP, KH-P) or highest likelihood (KH-L) trees. T-PTP tests were based on 100 permutations. KH-P, Kishino–Hasegawa test with parsimony; KH-L, Kishino–Hasegawa test with maximum likelihood; MT, missing taxa.

Table 3 Divergence times (Myr) based on 12S rRNA transversions

Divergence event	N	Mean	Standard deviation	Standard error
Among Paenungulates	3	54.8	4.2	2.4
Paenungulates to golden mole	3	67.1	8.7	5.0
Paenungulates to aardvark	3	74.0	12.0	6.9
Paenungulates to elephant shrew	3	79.9	9.9	5.7
African clade to other 13 orders	78	91.1	15.5	1.6

and that golden moles, elephant shrews, aardvarks and paenungulates are part of the same clade.

To corroborate these findings, we obtained sequences from exon 28 of the von Willebrand factor (vWF) gene for golden mole, hedgehog and pangolin. Adding these to the existing vWF data set⁹, we found high bootstrap support for the inclusion of golden moles with paenungulates, elephant shrews and aardvarks (Fig. 1b and Table 1). Sequences from the α -2B adrenergic receptor gene (A2AB) also support the association of golden moles with paenungulates, elephant shrews and aardvarks (Fig. 1c and Table 1). Parsimony and maximum-likelihood trees supporting the paenungulate–golden mole–aardvark–elephant shrew clade are significantly better than the best trees that constrain insectivore monophyly (Table 2).

This expanded clade, which includes five placental orders plus golden moles, has not been previously hypothesized on the basis of morphological or molecular data. Elephants, sirenians, hyraxes, golden moles, aardvarks and elephant shrews show a variety of ecological and morphological specializations and it is not surprising that morphology has not elucidated their common ancestry, now evident from DNA sequences. It is notable that all six of these groups are of probably African origin or, in the case of the aquatic sirenians, from along the margins of the former Tethys Sea^{5,6,13}. In two cases (golden moles and elephant shrews), geographic distribution has been restricted to Africa for the complete temporal range of these taxa⁵. Thus geographic evidence adds to the molecular data in support of this 'African origin' clade. The radiation of the African clade parallels endemic radiations of other vertebrate taxa on Southern Hemisphere continents during the breakup of Gondwanaland; for example, marsupials and passerine birds in Australia²⁰, and marsupials, edentates and notoungulates in South America⁵.

Paenungulate orders diverged from each other 51 to 59 million years (Myr) ago, as deduced from 12S rRNA transversions (Table 3). Deeper in the African clade, average divergence times between paenungulates and other lineages range from 67 to 80 Myr. The mean divergence time between taxa in the African clade and the other 13 orders of placental mammals is ~91 Myr. These divergence times support the hypothesis that many eutherian orders arose before the extinction of dinosaurs at the end of the Cretaceous²¹ and imply that conventional views on the origins of the African mammal fauna⁵ are incorrect. Africa's window of isolation extended from the Late Cretaceous, when Africa became separated from South America, to the early Cenozoic, when tenuous connections developed between northern Africa and Europe. The window of isolation extended from at least 80 Myr (ref. 20), if not earlier, until the early Cenozoic. The traditional view is that condylarths, prosimian primates and creodont carnivores reached Africa from the north after the docking of Africa with Europe⁵. From the condylarth stock, groups such as proboscideans and sirenians ostensibly originated in

Africa. Other elements of the African mammal fauna, including perissodactyls, artiodactyls, insectivores and living carnivore families, presumably arrived in the Neogene with the establishment of the Arabian Peninsula. Evidence for an extensive African clade, including taxa with divergence times as old as 80 Myr, is inconsistent with this view. The ancestor of the African clade probably resided in Africa before the window of isolation and did not arrive from the north in the early Cenozoic. The role of geographic isolation and continental break-up in the early diversification of placental mammals is potentially more important than previously recognized. □

Methods

Amplification and sequencing. 12S rRNA and tRNA genes were amplified and sequenced as described²². 16S rRNA genes were amplified using primers for valine-tRNA (for example, 5'-tacaccyaraagatttca-3') and leucine-tRNA (for example, 5'-agaggrtttgaaacctcg-3') and sequenced. Accession numbers for the new mitochondrial sequences (*Echymipera kalubu* (bandicoot); *Dromiciops gliroides* (monito del monte); *Sorex palustris* (shrew); *Manis* sp. (pangolin); *Amblysomus hottentotus* (golden mole); *Procavia capensis* (hyrax); *Trichechus manatus* (manatee); *Orycteropus afer* (aardvark); *Elephantulus rufescens* (elephant shrew)) are U97335–U97343. 12S rRNA sequences for several of these taxa have been deposited in GenBank (M95108 (golden mole), U61073 (monito del monte), U61079 (pangolin), U61083 (manatee), U61084 (hyrax)). Accession numbers for additional mitochondrial sequences are as follows: cow (J01394); blue whale (X72204); fin whale (X61145); horse (X79547); cat (U20753) harbour seal (X63726); grey seal (X72004); human (J01415); gorilla (D38114); orang-utan (D38115); guinea-pig (L35585); hedgehog (X88898); rat (X14848); mouse (J01420); opossum (Z29573); platypus (U33498; X83427). Exon 28 of the vWF gene was amplified and sequenced as described⁹. Accession numbers for *Manis* sp., *Erinaceus europaeus* (hedgehog), and *Amblysomus hottentotus* vWF sequences are U97534–U97536. Additional vWF sequences are from ref. 9. Part of the single-copy, intronless A2AB gene was amplified using the primers A2ABFOR (5'-ascctactcngtcgagcncacng-3') and A2ABREV (5'-ctgttgtagtagccatccaraaraaytg-3'). PCR products were cloned into a T/A cloning vector (Promega) and both strands were sequenced for at least two clones using the Thermo Sequenase fluorescent-labelled primer cycle sequencing kit (Amersham). Accession numbers for the new A2AB sequences (*Elephas maximus* (elephant); *Orycteropus afer* (aardvark); *Macroscelides proboscideus* (elephant shrew); *Amblysomus hottentotus* (golden mole); *Procavia capensis* (hyrax); *Erinaceus europaeus* (hedgehog); *Talpa europaea* (mole)) are Y12520–Y12526. Additional α -2 adrenergic sequences are M34041 (human); M32061 (rat); (L00974) (mouse), and U25722–U25724 (guinea-pig).

Sequence alignment and phylogenetic analysis. Sequences were aligned using CLUSTAL W (ref. 23). rRNA alignments were modified in view of secondary structure^{24,25}. Ambiguous regions were omitted from subsequent analyses²⁶; this resulted in 2,152, 1,261 and 1,132 nucleotide positions, respectively, for the mt, vWF and A2AB genes. The mt, vWF and A2AB data sets contain 810, 497 and 393 informative sites, respectively. Phylogenetic analyses (parsimony, minimum evolution with Tamura–Nei²⁷ and Logdet²⁸ distances, maximum likelihood under the HKY85 (ref. 29) model) were conducted with PAUP 4.0d52–54, written by D. L. Swofford. The mitochondrial tree was rooted using platypus and marsupial sequences. The vWF tree was rooted with the sloth⁷; alternatively, rooting with either hedgehog or rodents supports the 'African' clade and contradicts insectivore monophyly. For the A2AB tree, sequences with the suffix B are from the A2AB subfamily. GuineaPigA and GuineaPigC sequences are from other subfamilies in the α -2 adrenergic receptor family and were used as outgroups. Bootstrap analyses used full heuristic searches with 500 replications for parsimony and minimum

evolution and 100 replications for maximum likelihood. Shape parameters for the gamma distribution were estimated from minimum length trees²⁶ and were 0.32 (mtDNA), 0.59 (vWF) and 0.52 (A2AB).

Divergence times. 12S rRNA transversions accumulated linearly as far back as the eutherian–metatherian split²⁴. Nine independent cladogenic events were selected based on 12S rRNA sequence availability and paleostratigraphic data^{10,24,30} (for example, *Rattus* to *Mus* (14 Myr); *Sus* to *Tayassu* (45 Myr); ruminants to Cetacea (60 Myr); *Erinaceus* to Metatheria (130 Myr)). Relative rates were calculated in reference to xenarthrans. Tamura–Nei transversion distances (transversions only) were adjusted for relative rate differences³⁰ against the xenarthran standard. Rate-adjusted estimates of sequence divergence were regressed against paleostratigraphic divergence estimates for each of the nine calibration points (origin forced through zero; $r^2 = 0.97$; $P = 0.0000002$). The resulting equation (divergence time (in Myr) = sequence divergence/0.00063) was used to estimate interordinal divergence times after making similar adjustments for relative rates. Additional details will be presented elsewhere (M.S., manuscript in preparation).

Received 23 December 1996; accepted 18 April 1997.

- Novacek, M. J. Mammalian phylogeny: shaking the tree. *Nature* **356**, 121–125 (1992).
- Huxley, T. H. On the application of the laws of evolution to the arrangement of the Vertebrata, and more particularly to the Mammalia. *Proc. R. Soc. Lond.* **43**, 649–662 (1880).
- Matthew, W. D. The Carnivora and Insectivora of the Bridger Basin, Middle Eocene. *Mem. Am. Mus. Nat. Hist.* **9**, 291–567 (1909).
- MacPhee, R. D. E. & Novacek, M. J. in *Mammal Phylogeny Vol. 2, Placentals* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 13–31 (Springer, New York, 1993).
- Carroll, R. L. *Vertebrate Paleontology and Evolution* (Freeman, New York, 1988).
- Gheerbrant, E., Sudre, J. & Cappetta, H. A Palaeocene proboscidean from Morocco. *Nature* **383**, 68–70 (1996).
- Lavergne, A., Douzery, E., Stiehler, T., Catzeflis, F. M. & Springer, M. S. Interordinal mammalian relationships: evidence for paenungulate monophyly is provided by complete mitochondrial 12S rRNA sequences. *Mol. Phylog. Evol.* **6**, 245–258 (1996).
- Madsen, O., Deen, P. M. T., Pesole, G., Saccone, C. & de Jong, W. W. Molecular evolution of mammalian aquaporin-2: further evidence that elephant shrew and aardvark join the paenungulate clade. *Mol. Biol. Evol.* **14**, 363–371 (1997).
- Porter, C. A., Goodman, M. & Stanhope, M. J. Evidence on mammalian phylogeny from sequences of exon 28 of the von Willebrand factor gene. *Mol. Phys. Evol.* **5**, 89–101 (1996).
- Novacek, M. J. et al. Mammalian evolution and the interphotoreceptor retinoid binding protein (IRBP) gene: convincing evidence for several superordinal clades. *J. Mol. Evol.* **43**, 83–92 (1996).
- Cope, E. D. The condylarthra. *Am. Nat.* **18**, 790–805, 892–906 (1884).
- Fischer, M. S. & Tassy, P. in *Mammal Phylogeny Vol. 2, Placentals* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 217–234 (Springer, New York, 1993).
- de Jong, W. W., Zweers, A. & Goodman, M. Relationship of aardvark to elephants, hyraxes and sea cows from α -crystallin sequences. *Nature* **292**, 538–540 (1981).
- de Jong, W. W., Leunissen, J. A. M. & Wistow, G. J. in *Mammal Phylogeny Vol. 2, Placentals* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 5–12 (Springer, New York, 1993).
- Faith, D. P. Cladistic permutation tests for monophyly and nonmonophyly. *Syst. Zool.* **40**, 366–375 (1991).
- Kishino, H. & Hasegawa, M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* **29**, 170–179 (1989).
- Glover, T. D. Aspects of sperm production in some east African mammals. *J. Reprod. Fertil.* **35**, 45–53 (1973).
- Hartenberger, J. L. Hypothese paleontologique sur l'origine des Macroscelidea (Mammalia). *C.R. Acad. Sci.* **302**, 247–249 (1986).
- Novacek, M. in *Macromolecular Sequences in Systematic and Evolutionary Biology* (ed. Goodman, M.) 3–41 (Plenum, New York, 1982).
- Sibley, C. G. & Ahlquist, J. E. Reconstructing bird phylogeny by comparing DNAs. *Sci. Am.* **254**, 82–92 (1986).
- Hedges, S. B., Parker, P. H., Sibley, C. G. & Kumar, S. Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381**, 226–229 (1996).
- Springer, M. S., Hollar, L. J. & Burk, A. Compensatory substitutions and the evolution of the mitochondrial 12S rRNA gene in mammals. *Mol. Biol. Evol.* **12**, 1138–1150 (1995).
- Thompson, J. D., Higgins, G. D. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
- Springer, M. S. & Douzery, D. Secondary structure and patterns of evolution among mammalian mitochondrial 12S rRNA molecules. *J. Mol. Evol.* **43**, 357–373 (1996).
- De Rijk, P., Van de Peer, Y., Chapelle, S. & De Wachter, R. *Nucleic Acids Res.* **22**, 3495–3501 (1994).
- Swofford, D. L., Olsen, G. J., Waddell, P. J. & Hillis, D. M. in *Molecular Systematics* (eds Hillis, D. M., Moritz, C. & Mable, B. K.) 407–514 (Sinauer, Sunderland, MA, 1996).
- Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**, 512–526 (1993).
- Lockhart, P. J., Steel, M. A., Hendy, M. D. & Penny, D. Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* **11**, 605–612 (1994).
- Hasegawa, M., Kishino, H. & Yano, T. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **21**, 160–174 (1985).
- Arnason, U., Gullberg, A., Janke, A. & Xu, X. Pattern and timing of evolutionary divergences among hominoids based on analyses of complete mtDNAs. *J. Mol. Evol.* **43**, 650–661 (1996).

Acknowledgements. This work was supported by the Alfred P. Sloan Foundation, the European Commission, the NSF, the Nuffield Foundation and the Royal Society. We thank D. Willemsen for technical assistance, D. Swofford for permission to use PAUP 4.0d52-54, and F. Catzeflis, E. Harley, J. Kirsch, G. Olbricht, J. Wensing and the Noorder Zoo for tissue samples.

Correspondence and requests for materials should be addressed to M.S.S. (e-mail: mark.springer@ucr.edu).

Hypothermia in foraging king penguins

Y. Handrich*, R. M. Bevan†, J.-B. Charrassin*, P. J. Butler†, K. Pütz‡, A. J. Woakes†, J. Lage* & Y. Le Maho*

* Centre d'Ecologie et Physiologie Énergétiques, Centre National de la Recherche Scientifique, 23 rue Becquerel, 67087 Strasbourg cedex 2, France

† School of Biological Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

‡ Abteilung Meereszoologie, Institut für Meereskunde, Düsterbrook Weg 20, D-24105 Kiel, Germany

The ability to dive for long periods increases with body size¹, but relative to the best human divers, marine birds and mammals of similar or even smaller size are outstanding performers. Most trained human divers can reach a little over 100 m in a single-breath dive lasting for 4 min (ref. 2), but king and emperor penguins (weighing about 12 and 30 kg, respectively) can dive to depths of 304 and 534 m for as long as 7.5 and 15.8 min, respectively^{3–5}. On the basis of their assumed metabolic rates, up to half of the dive durations were believed to exceed the aerobic dive limit, which is the time of submergence before all the oxygen stored in the body has been used up^{4,6,7}. But in penguins and many diving mammals^{7,8}, the short surface intervals between dives are not consistent with the recovery times associated with a switch to anaerobic metabolism⁴. We show here that the abdominal temperature of king penguins may fall to as low as 11 °C during sustained deep diving. As these temperatures may be 10 to 20 °C below stomach temperature, cold ingested food cannot be the only cause of abdominal cooling. Thus, the slower metabolism of cooler tissues resulting from physiological adjustments associated with diving *per se*, could at least partly explain why penguins and possibly marine mammals can dive for such long durations.

King penguins are pelagic predators. To obtain food for their chicks, the parents forage at sea up to the subantarctic or polar frontal zones, 300–1,000 km away from their breeding colony^{9,10}. They essentially rely on myctophid fish, of which most are captured in daytime at 150–300 m depths¹¹. As sea temperatures there are 4 °C or lower, their stomachs are cooled by ingested prey^{12,13}. In freely foraging king penguins, which normally have a body temperature of 38 °C on land, stomach temperatures as low as 19 °C have been reported^{11,14}. There is a 2–4 °C fall in body temperature during free diving activity in seals^{15,16} and birds^{17–19} and it has been suggested that a slight reduction in body temperature during diving might enhance aerobic diving time^{6,15,17,18}. The cold food that antarctic animals eat could contribute to this hypothermia^{14,20–22}, or the aerobic dive limit (ADL) of penguins might be prolonged by a process of temperature-induced metabolic suppression that is independent of stomach cooling.

To investigate these possibilities, the separate influences of feeding and diving on the abdominal temperatures of foraging animals have to be determined. It is important to obtain simultaneous measurements of the temperatures inside and outside the stomach while the animals are freely diving, so that the extent of the temperature changes in relation to diving and feeding activity during the course of a foraging trip can be found. We therefore implanted three data loggers into each of 12 free-ranging king penguins (see Methods). The data loggers (Fig. 1) measured the temperature of each bird at the top (T_{abtop}) and bottom (T_{abbot}) of the abdomen, as well as inside the stomach (T_{stom}). Hydrostatic pressure was also recorded to monitor the diving behaviour. Both the upper-abdominal and stomach loggers measured the full range of temperatures; the lower-abdominal device recorded temperatures

only below 22.7 °C (see Methods and Fig. 2). Before the birds went to sea, upper-abdominal and stomach temperatures did not differ significantly, remaining at the mean value ($\pm$ s.d.) of 37.5 ± 0.4 °C ($n = 8$ birds). Once the birds were at sea, the stomach temperature was significantly higher between diving bouts than when the birds were ashore (38.3 ± 0.4 °C; $n = 8$), and systematically higher (0.5 ± 0.03 °C; $n = 7$, $P < 0.05$) than the upper-abdomen temperature. In contrast, both upper-abdominal and stomach temperatures fell during foraging dives. The lowest temperature recorded during a bout of deep diving ($n = 52$ for dives deeper than 50 m) ranged from 18.2–37.0 °C in the stomach and 25.8–37.1 °C in the upper abdomen ($n = 7$ birds). At any given time, the difference between these two temperatures could be as much as 12 °C. Surprisingly, this difference could be either positive or negative (Fig. 3A). Furthermore, T_{abbot} was also recorded (that is, it fell below 22.7 °C) during prolonged (>3 h) deep diving (Fig. 2). From ten data sets available, these very low temperatures occurred during 11% of the time of the overall foraging trip in 7 of the 10 penguins, with 83% of the total duration of these events being during the last two days of foraging. When recorded, temperature in the lower abdomen was always lower than the other two temperatures, the difference ranging between 2.9 and 20.1 °C. It was at least 2 °C below stomach temperature. Thus, whereas most previous studies have attributed the decreases in either stomach^{14,20} or abdominal^{21,22} temperatures to the ingestion of cold prey, our results show that the lower abdomen

is cooled not just by the contents of the stomach but also by other mechanisms (Fig. 3A).

The stomach of the king penguin is divided into two distinct regions, each with a different morphology and thermal insulation (Fig. 1). The lower part, the gizzard, is muscular, well perfused, and well insulated from the abdominal cavity by a thick (1–2 cm) layer of abdominal fat. In contrast, the upper part, the proventriculus, has a thin wall and is in direct contact with the upper abdominal cavity. At the start of a foraging trip, the gizzard is empty, so any ingestion is easily detected¹³; it can be assumed that the stomach temperature at this time is that of the whole stomach and that there is no substantial temperature gradient within the stomach (Fig. 3B). The temperature in the upper abdomen then decreases by ~ 1 °C during the first bout of shallow diving, although there is no evidence of any ingestion (event a in Fig. 3B).

A similarly slight temperature decrease has been reported in the dorsal aortic blood of a freely diving Weddell seal¹⁵ (*Leptonychotes weddellii*), which was explained by a reflex decrease in the perfusion of peripheral and/or inactive muscle tissues that caused a reduction in metabolic rate¹⁵. This idea is supported by data from prolonged deep-diving bouts performed by king penguins. At this stage, a change in stomach temperature was observed for all diving events. This phenomenon has been previously described¹³ and was attributed to the movement of the stomach recorder through different strata of food at different temperatures. However, the temperatures

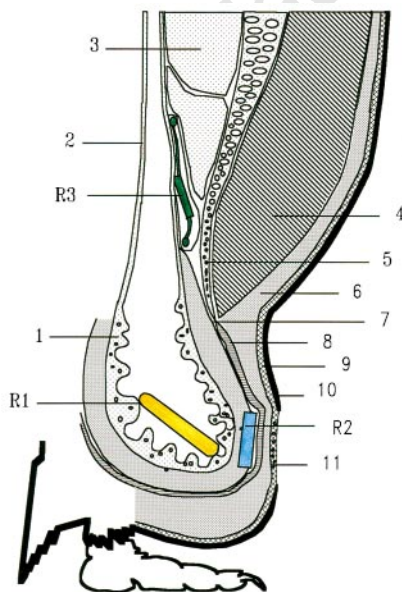


Figure 1 Sagittal view of the abdominal region of a king penguin showing the location of three recorders: R1, time/temperature recorder located in the stomach; R2, time/temperature/depth recorder implanted in the abdominal fat of the lower abdomen, with an external thermistor placed against the stomach wall; R3, time/temperature/heart rate recorder implanted under the sternum and against the stomach wall in the upper abdomen. Parts of the penguin's body indicated are: (1) muscular wall of the gizzard; (2) thin wall of the proventriculus; (3) liver lobes; (4) pectoral muscles; (5) sternum; (6) subcutaneous adipose layer; (7) abdominal and visceral fat deposits; (8) external and internal oblique muscles; (9) skin; (10) feathers; (11) well perfused brood patch.

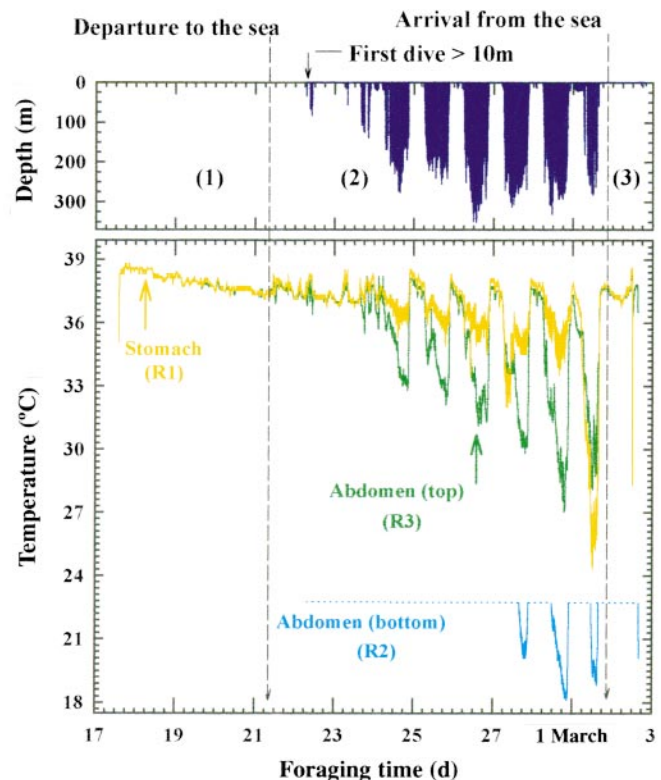


Figure 2 Typical data obtained from a single bird corresponding to: (1) the end of the brooding shift (range, 5–10 d; mean $\pm$ s.d.: 7.3 ± 1.9 , $n = 12$); (2) the foraging trip (range 5.1–10.2 d; mean, 7.5 ± 1.6); (3) the beginning of the following brooding shift (less than 36 h). The typically gradual decrease in stomach temperature (T_{stom}) is shown over the 2–5 d following implantation, as well as the further periodic decreases in both T_{stom} and upper-abdominal temperature (T_{abtop}) which were synchronized with diving bouts. In the example shown here, the lower-abdominal temperature (T_{abbot}) fell below 22.7 °C (the upper limit of the range of R2) in the three last diving bouts, reaching values as low as 18.1 °C.

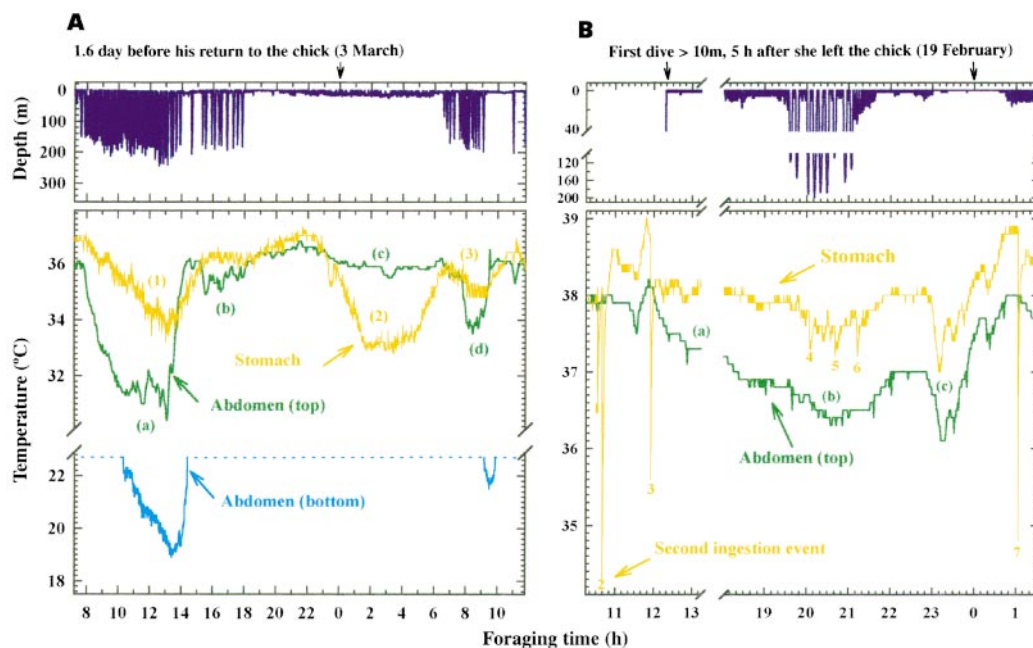


Figure 3 A, Representative changes of the three recorded temperatures during different diving activity two days before the end of the foraging trip (bird N departed on 21 February at 08:02 h for 10.5 d of foraging), showing the temporal changes in temperature that occur inside the abdominal cavity. At any given time, temperatures could differ by more than 15 °C. There is a relationship between the reduction in T_{abtop} and diving behaviour, with slight decreases during periods of shallow diving (**c**) and larger decreases during deep diving bouts (**a**, **b**, **d**). Although T_{abtop} may be several degrees lower than T_{stom} (1 is for T_{stom} , a is for T_{abtop}) the source of cooling could still be the ingested food because large thermal differences exist in the stomach (see text). Note that, during the resting phase preceding shallow dives, T_{stom} is still higher than T_{abtop} , which might indicate local heat production during digestion. When temperature data are available for the bottom part of the abdomen (when T_{abbot} is below 22.7 °C) as shown here in two

diving bouts, these events roughly parallel T_{abtop} , with a slight delay, even during the rewarming phase. **B**, Representative changes in body temperature of a king penguin at sea 5 h after leaving the chick (bird K departed on 18 February at 06:49 h for 8.0 d of foraging). At this stage, the stomach is empty and any ingestion of food or water is detected by a typical precipitous fall in T_{stom} (ingestion events 1 to 7; ref. 13). Although there was no ingestion during the 8 h preceding the first deep-diving bout, T_{abtop} progressively decreased by more than 1 °C, suggesting that this decline is due to diving *per se*. The further decrease of T_{abtop} during a deep diving bout (event **b**) could be partly due to ingestion (events 4–6). However, the transient drop in both T_{stom} and T_{abtop} preceding the rewarming phase (event **c**), which routinely occurs during a surface phase, argues for a mechanism independent of feeding.

outside the stomach show similar changes (Fig. 4) and may instead be attributed to reflex adjustments in blood perfusion that are associated with diving *per se*^{23,24}.

The cooling of less well perfused tissues during descent and the redistribution of blood of different temperatures during ascent²³ would explain these observations. This means that the monitoring of stomach temperature in marine birds, which has proved useful for detecting prey ingestion events^{12,13}, should be re-evaluated in terms of diving hypothermia for calculating the mass of prey ingested by diving species.

As the diving bouts continue, the stomach temperature recorder (located in the gizzard¹³) is progressively surrounded by more food, and the sensor becomes unable to detect single ingestion events¹³. When sufficient food has been ingested to fill the proventriculus, a thermal gradient may develop between this cooled compartment and the warmed gizzard²⁵. At this stage, the upper abdomen close to the proventriculus was often at a lower temperature than the stomach and this could, at least partly, be attributed to cooling from the proventriculus (event a in Fig. 3A). Ingested food could then contribute to the cooling of the core temperature close to the liver, while the digestive (thermogenic) process continues in the lower gizzard. As this lower part of the stomach is well insulated, the temperature here can be 20 °C higher than that in the lower abdomen, even though the sensors are only 5 cm apart. Towards the end of the foraging trips, however, the temperature in the stomach may progressively decrease and become less than that in the lower

abdomen (Fig. 2, event 2 and 3A). This could result from a slowing down in the processes of digestion and assimilation by the foraging penguin, and from the storage of cold food for the chick.

The cause of the decline in body temperature in freely diving endotherms has been debated for over fifty years^{1,15,23,26}; it may simply be the result of an increase in blood flow to the periphery which causes increased heat loss^{1,26}, or there may be a local metabolic depression as a result of reduced blood perfusion^{15,23}. These explanations may not be mutually exclusive, at least for breeding king penguins at the beginning of a diving bout. Our results are consistent with the possibility that the king penguin is engaged in an active process of temperature reduction, with the well vascularized brood patch and flippers being effective surfaces for heat loss (Fig. 1). When a diving bout ends and the deep regions of the body gradually rewarm, blood from the active pectoral muscles may be redistributed preferentially to the deeper tissues in order to reduce heat loss from the periphery and to rewarm the core²³. The question is, would the energy savings from such a regional hypothermia be of overall benefit, considering the cost of rewarming the body? The fact that rewarming occurs while the penguins continue to dive (Figs 2, 4) suggests that the 'waste' heat of exercise²⁷, rather than shivering thermogenesis, may be used for rewarming. Foraging birds must also travel between prey patches as well as back to the breeding site. 'Waste' heat is therefore readily available during foraging activity and rewarming could occur at little or no extra cost to the birds.

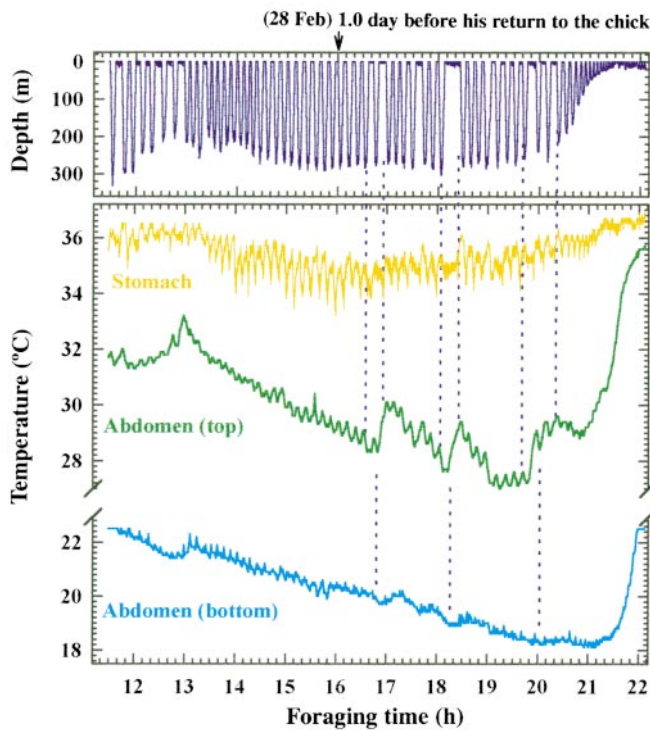


Figure 4 Short-term variations of the three abdominal temperatures near the end of a deep-diving bout and 1 d before the end of the foraging trip (bird O returns to the chick on 01 March at 16:08h after 8.2 d at sea). There are systematic oscillations of the three temperatures which are clearly associated with individual dives. The synchronization of the three temperatures with diving is not precise, which could be due to physical and electronic delays (sampling interval). Long interdiving intervals indicate a tendency for temperatures (particularly T_{abtop}) to fall during a dive and to increase during the time spent at the surface. In all cases, the progressive decrease in temperature is probably due to the fact that the post-dive interval is not long enough to allow for a full recovery. These oscillations and progressive decreases may be due to a diving-induced redistribution of blood at different temperatures from different regions of the body.

Our results show that during deep dives, temperatures in certain body regions of freely foraging penguins can decrease much more dramatically ($>10^{\circ}\text{C}$) than in the stomach, which is cooled predominantly by the ingestion of cold prey. These temperature decreases, which would lead to a metabolic depression, might help explain the extraordinary diving performance of king penguins and other marine endotherms. The extent of the metabolic depression caused by the lowered temperatures in diving king penguins might confer upon them a significant overall energetic benefit during their foraging trip, which would be a particular advantage at the time breeding parents need to minimize their own energy expenditure while accumulating food in the stomach for their chicks. This energy saving would also be analogous to that which results from torpid periods in hibernators, despite the cost of arousal^{28,29}. Finally, although they are much poorer at diving than penguins, the best human divers rely on voluntary control of autonomic processes by using techniques such as yoga²; these may induce hypometabolism³⁰, which, according to our results, is a key process for tolerating an extended breath-hold.

Methods

Twelve birds (males and females) were equipped two days after the start of their first brooding shift. All chicks were less than 12 d old. The stomach recorder (R1 in Fig. 1) was swallowed before implantation of the lower- and upper-abdominal temperature recorders. R2 and R3, respectively, were implanted under halothane anaesthesia and sutured in place. For R1:

dimensions (d), 105 mm $\times$ 16 mm diameter¹³; temperature range (R), 16 to 41°C ; sampling interval (SI), 16 s. For R2: $d = 64 \text{ mm} \times 38 \text{ mm} \times 15 \text{ mm}$ (Wildlife Computer, Inc.); R , -2.5 to 22.7°C ; SI, 30 s; depth range, 0–500 m; SI, 3 s. For R3: $d = 55 \text{ mm} \times 24 \text{ mm} \times 6 \text{ mm}$; R , 12 – 41°C ; SI, 60 s. Sampling of stomach temperature (T_{stom}) started immediately; upper abdominal temperature (T_{abtop}) and hydrostatic pressure sampling started after the first foraging dive that exceeded 10 m. Pressure was recorded from 11 birds. Bottom abdominal temperature (T_{abbot}) was from 10 birds; T_{abtop} was from 8 birds; and T_{stom} was from 8 birds. Data from all four channels were recorded simultaneously in 7 birds. The 12 birds were recaptured 1–2 days after returning to their chicks. Only one failed in brooding after removal of the three recorders. R2 and R3 were removed surgically and R1 by stomach flushing. All thermistors were simultaneously calibrated ($\pm 0.2^{\circ}\text{C}$) after the experiment.

Received 11 December 1996; accepted 18 April 1997.

- Kooyman, G. L. *Diverse Divers* (ed. Farner, D. S.) 1–200 (Springer, Berlin, 1989).
- Corriol, J. H. *La Plongée en Apnée. Physiologie et Médecine* 2nd edn (Masson, Paris, 1996).
- Kooyman, G. L. *et al.* Diving behavior and energetics during foraging cycles in king penguins. *Ecol. Monogr.* **62**, 143–163 (1992).
- Kooyman, G. L. & Kooyman, T. G. Diving behavior of emperor penguins nurturing chicks at Coulman island, Antarctica. *Condor* **97**, 536–549 (1995).
- Ancel, A. *et al.* Foraging behaviour of emperor penguins as a resource detector in winter and summer. *Nature* **360**, 336–338 (1992).
- Culik, B. M. *et al.* Diving energetics in king penguins (*Aptenodytes patagonicus*). *J. Exp. Biol.* **199**, 973–983 (1996).
- Kooyman, G. L. *et al.* Heart rates and swim speeds of emperor penguins diving under sea ice. *J. Exp. Biol.* **165**, 161–180 (1992).
- Boyd, I. L. & Croxall, J. P. Dive durations in pinnipeds and seabirds. *Can. J. Zool.* **74**, 1696–1705 (1996).
- Guinet, C. *et al.* Foraging behaviour of satellite-tracked king penguins in relation to the sea surface temperature obtained by satellite telemetry at Crozet Archipelago during the austral summer: a three year study. *Mar. Ecol. Prog. Ser.* (in the press).
- Bost, C. A. *et al.* Foraging habitat and food intake of satellite-tracked king penguins during the austral summer at Crozet Archipelago. *Mar. Ecol. Prog. Ser.* (in the press).
- Pütz, K. & Bost, C.-A. Feeding behavior of free-ranging king penguins (*Aptenodytes patagonicus*). *Ecology* **75**, 489–497 (1994).
- Wilson, R. P. & Culik, B. M. The cost of a hot meal: facultative specific dynamic action may ensure temperature homeostasis in post-ingestive endotherms. *Comp. Biochem. Physiol.* **100A**, 151–154 (1991).
- Wilson, R. P. *et al.* Reliability of stomach temperature changes in determining feeding characteristics of seabirds. *J. Exp. Biol.* **198**, 1115–1135 (1995).
- Pütz, K. Aspects of the feeding ecology of emperor penguins (*Aptenodytes forsteri*) and king penguins (*Aptenodytes patagonicus*). *Rep. Polar Res.* **136**, 1–139 (1994).
- Hill, D. R. *et al.* Heart rate and body temperature during free diving of Weddell seals. *Am. J. Physiol.* **253**, R344–R351 (1987).
- Ponganis, P. J., Kooyman, G. L., Castellini, M. A., Ponganis, E. P. & Ponganis, K. V. Muscle temperature and swim velocity profiles during diving in a Weddell seal, *Leptonychotes weddellii*. *J. Exp. Biol.* **183**, 341–348 (1993).
- Culik, B. M. *et al.* Core temperature variability in diving king penguins (*Aptenodytes patagonicus*): a preliminary analysis. *Polar Biol.* **16**, 371–378 (1996).
- Butler, P. J., Bevan, R. M., Woakes, A. J., Croxall, J. P. & Boyd, I. L. The use of data loggers to determine the energetics and physiology of aquatic birds and mammals. *Braz. J. Med. Biol. Res.* **28**, 1307–1317 (1995).
- Bevan, R. M. *et al.* Heart rate and abdominal temperatures of free-ranging South Georgian shags, *Phalacrocorax georgianus*. *J. Exp. Biol.* **200**, 661–675 (1997).
- Wilson, R. P., Cooper, J. & Plötz, J. Can we determine when marine endotherms feed? *J. Exp. Biol.* **167**, 267–275 (1992).
- Woakes, A. J., Butler, P. J., Bevan, R. M. & Boyd, I. L. The metabolic rate of free ranging antarctic animals. *Biotelemetry XII* **21**, 3–223 (1992).
- Woakes, A. J., Butler, P. J. & Bevan, R. M. Implantable data logging system for heart rate and body temperature: its application to the estimation of field metabolic rates in Antarctic predators. *Med. Biol. Eng. Comput.* **33**, 145–151 (1995).
- Scholander, P. F., Irving, L. & Grinnel, S. W. On the temperature and metabolism of the seal during diving. *J. Cell Comp. Physiol.* **19**, 67–78 (1942).
- Bevan, R. M. & Butler, P. J. Cardiac output and blood flow distribution during swimming and voluntary diving of the tufted duck (*Aythya fuligula*). *J. Exp. Biol.* **168**, 199–217 (1992).
- Carey, F. G. *A Companion to Animal Physiology* (eds Taylor, C. R., Johnsen, K. & Bolis, L.) 216–235 (Cambridge Univ. Press, 1982).
- Kooyman, G. L., Wahrenbrock, E. A., Castellini, M. A., Davis, R. W. & Sinnett, E. E. Aerobic and anaerobic metabolism during voluntary diving in Weddell seals: evidence of preferred pathways from blood chemistry and behavior. *J. Comp. Physiol.* **138**, 335–346 (1980).
- Kettersen, E. D. & King, J. R. Metabolic and behavioral responses to fasting in the white-crowned sparrow (*Zonotrichia leucophrys gambelli*). *Physiol. Zool.* **50**, 115–129 (1977).
- Heldmaier, G., Steiger, R. & Ruf, T. *Life in the Cold. Ecological, Physiological and Molecular Mechanisms* (eds Carey, C., Florant, G. L., Wunder, B. A. & Horwitz, B.) 545–548 (Westview, Boulder, 1993).
- Ruf, T. & Heldmaier, G. The impact of daily torpor on energy requirements in the Djungarian hamster, *Phodopus sungorus*. *Physiol. Zool.* **65**, 994–1010 (1992).
- Heller, H. C., Elsner, R. & Rao, N. Voluntary hypometabolism in an Indian yogi. *J. Therm. Biol.* **12**, 171–173 (1987).

Acknowledgements. We thank G. Froget, Y. Ropert-Coudert, J.-N. Clech and all the 1995 winter team in Crozet for their assistance in the field, and A. Ancel, C.-A. Bost, B. M. Culik and A. Malan for comments on the manuscript. After approval by the ethics committee of the Institut Français pour la Recherche et la Technologie Polaires, this study was supported by a grant from this Institute and by an NERC Small Research grant.

Correspondence and requests for materials should be addressed to Y.H. (e-mail: yves.handrich@c-strasbourg.fr).

Colour tuning in human visual cortex measured with functional magnetic resonance imaging

Stephen Engel*, Xuemei Zhang & Brian Wandell

Stanford University, Department of Psychology, Building 420, Stanford, California 94305, USA

The primate retina contains three classes of cones, the L, M and S cones, which respond preferentially to long-, middle- and short-wavelength visible light, respectively. Colour appearance results from neural processing of these cone signals within the retina and the brain. Perceptual experiments have identified three types of neural pathways that represent colour: a red–green pathway that signals differences between L- and M-cone responses; a blue–yellow pathway that signals differences between S-cone responses and a sum of L- and M-cone responses; and a luminance pathway that signals a sum of L- and M-cone responses^{1–3}. It might be expected that there are neurons in the primary visual cortex with response properties that resemble these three perceptual pathways, but attempts to find them have led to inconsistent results^{4–7}. We have therefore used functional magnetic resonance imaging (fMRI) to examine responses in the human brain to a large number of colours. In visual cortical areas V1 and V2, the strongest response is to red–green stimuli, and much of this activity is from neurons receiving opposing inputs from L and M cones. A strong response is also seen with blue–yellow stimuli, and this response declines rapidly as the temporal frequency of the stimulus is increased. These responses resemble psychophysical measurements, suggesting that colour signals relevant for perception are encoded in a large population of neurons in areas V1 and V2.

Midget retinal ganglion cells and parvocellular neurons in the lateral geniculate nucleus (LGN) combine cone signals of opposite sign, as expected in chromatic pathways^{8–10}. Several of the quantitative colour properties of neurons in the retina and LGN, however, do not match perceptual measurements. For example, perceptual spatial resolution is highest for luminance patterns, but the retinal mosaic with the highest spatial resolution, the midget ganglion cells, receive opposing input from the long-wavelength-sensitive (L) cones and the middle-wavelength-sensitive (M) cones, as expected in a chromatic pathway^{11,12}.

Such discrepancies have focused attention on the cortex in an attempt to find neurons with properties that more closely resemble the three perceptual pathways^{7,10}. However, single-unit colour-tuning measurements in cortical area V1 have yielded conflicting results. Some measurements identify three populations of neurons, associated with particular cytochrome-oxidase blobs, that may correspond to the chromatic pathways^{5,13}, but other results show broad distributions of colour properties, and yet others report that most V1 neurons respond best to luminance rather than chromatic patterns^{4,6,7}.

To clarify the relationship between cortical activity and the perceptual colour pathways, we measured cortical colour tuning in areas V1 and V2 by using functional magnetic resonance imaging (fMRI). We used stimuli that reveal red–green and blue–yellow chromatic pathways in psychophysical experiments^{3,14,15}, and measured colour tuning by using contrast-reversing ‘checkerboard’ patterns of many colours and contrasts. In each experiment,

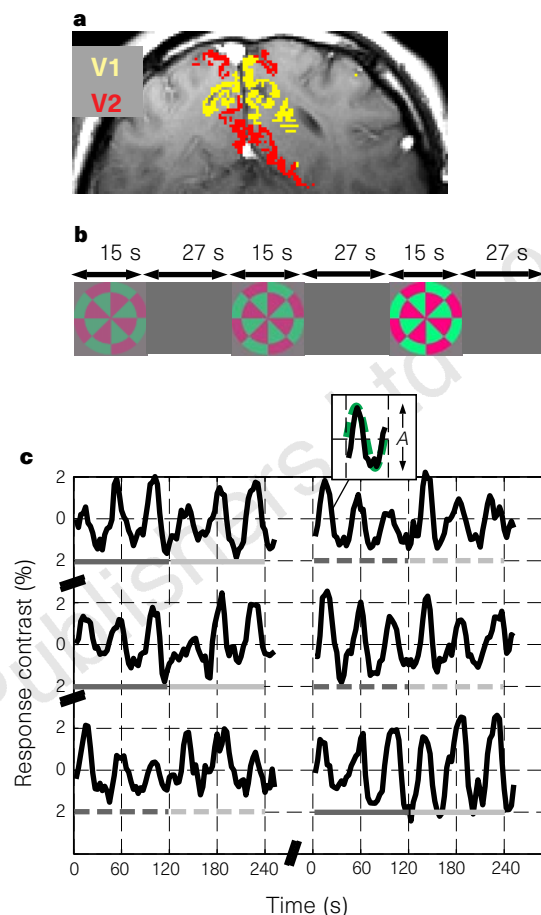


Figure 1 Experimental protocol. **a**, Cortical areas V1 (yellow) and V2 (red) (for observer B.W.). **b**, Stimulus time course (see Methods); only the central portion of the stimulus is shown (the actual stimulus contained many more checks). **c**, Typical time series of the fMRI signal (area V1, observer B.W., 4 Hz, red–green plane). Six stimuli were presented during each fMRI scan. The lines beneath the time series are shaded to indicate stimuli with a common colour direction; solid and dotted lines denote increasing and decreasing contrast series, respectively. Inset shows response amplitudes (A) to each stimulus, calculated by fitting harmonic functions to the time series.

observers viewed stimuli in six different colour directions at three levels of colour contrast (see Fig. 1). Each stimulus was presented for 15 s and was followed by a mean-field presentation for 27 s; we call this composite presentation a stimulus period. The amplitude of the harmonic component of the fMRI signal with the stimulus period (42 s) served to measure the response of the neural population to each stimulus. This response increased monotonically with stimulus contrast. Each experiment was conducted with contrast-reversal temporal frequencies of 1, 4 and 10 Hz.

The first experiments measured colour tuning in a red–green stimulus plane in which checkerboards contained various levels of L- and M-cone contrast and zero S-cone contrast. Results for two observers are shown in Fig. 2. The solid iso-response contours show the L- and M-cone contrasts of stimuli that produce a criterion response amplitude. The dashed contours represent an 80% confidence interval surrounding the iso-response contour.

These data show that a large proportion of the response to chromatic stimuli arises from neurons that receive opposing L- and M-cone signals. There are several ways to see the large contribution from such cells. First, the long straight edges of the iso-response contours fall along lines described by the equation $|L - M| = a \text{ constant}$, where L and M are the L- and M-cone

* Present address: Department of Psychology, UCLA, Franz Hall, Los Angeles, California 90095, USA.

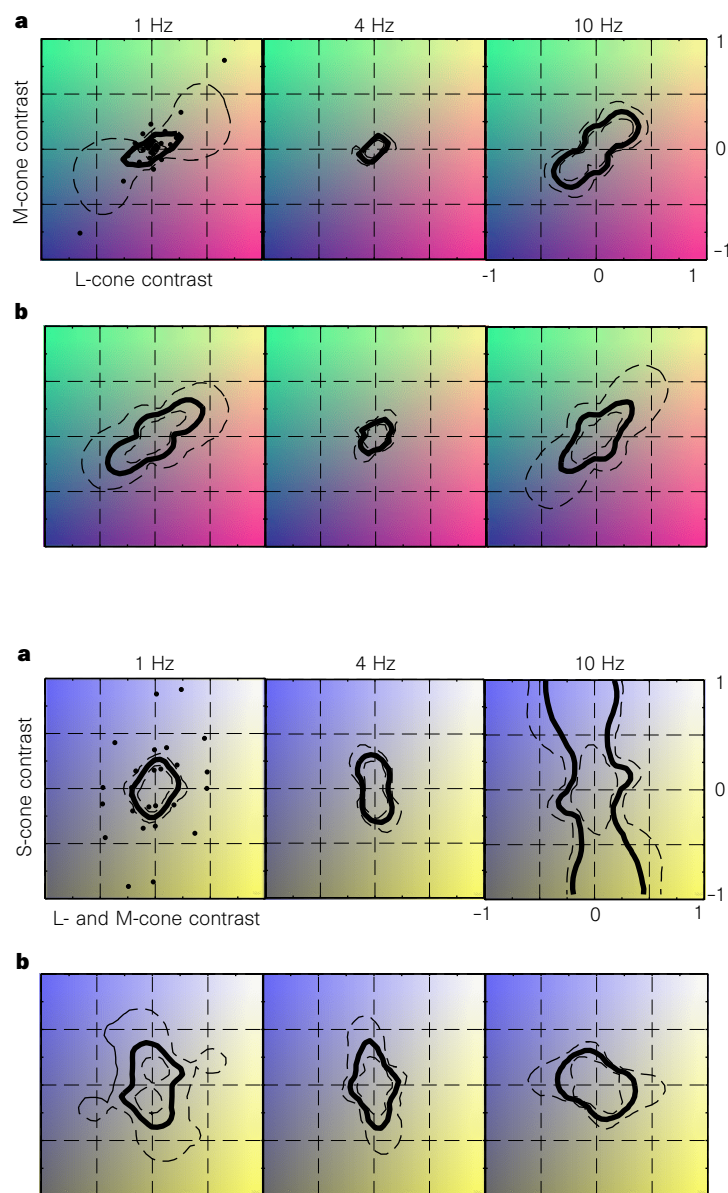


Figure 2 Colour tuning in a red–green plane measured in human cortical area V1. The solid contour shows the amount of cone contrast required to generate an fMRI response criterion level. The dotted contours represent 10% and 90% confidence intervals. The dots on the left in **a** show the L- and M-cone contrasts of the 18 different stimuli used in this experiment. The background colour shows the general appearance of the stimuli. The colour stimuli were presented at three temporal frequencies. Iso-response contours are shown for observer B.W. (**a**) and S.E. (**b**) measured using temporal frequencies of 1, 4 and 10 Hz. For both observers, the response levels represented by the iso-response contours are about twice as high at 4 Hz as at 1 Hz or 10 Hz.

Figure 3 Colour tuning in a blue–yellow plane measured in human cortical area V2. The horizontal axis represents equal amounts of L- and M-cone contrast. The vertical axis represents S-cone contrast. Iso-response contours (solid curves) for observer B.W. (**a**) and S.E. (**b**) at 1, 4 and 10 Hz temporal frequencies. For both observers, the response levels represented by the iso-response contours are about twice as high at 4 Hz as at 1 Hz or 10 Hz. Other features are as Fig. 2.

contrasts, respectively. Second, at all three temporal frequencies, more contrast is needed to obtain a given response level to L and M signals of common sign than of opposite sign; at 4 Hz, for example, a cone contrast of roughly $(0.05, -0.05, 0)$ yields the same response as a cone contrast of $(0.2, 0.2, 0)$. Hence, for this spatio-temporal pattern, cone contrast combined in opposite sign (a red–green chromatic stimulus) is four times more effective than cone signals combined in common sign (a luminance stimulus). Third, adding M-cone stimulation to a pure L-cone signal with the same sign reduces the fMRI response, creating a stimulus that falls inside the iso-response contour. Colour tuning measured in area V2 was virtually the same as colour tuning in area V1; the iso-response contours from V2 fell almost entirely within the confidence intervals obtained from area V1.

A second series of experiments measured colour tuning in a blue–yellow plane, with stimulus axes specified in terms of S-cone contrast and combined equal amounts of L- and M-cone contrast. Iso-response contours in this plane from area V2 are shown in Fig. 3, and responses in V1 were very similar. At low temporal frequencies, we again obtained a robust fMRI signal in response to chromatic stimulation; responses in the blue–yellow $(1, 1, -1)$ colour direction were equal to or greater than responses in the luminance

direction $(1, 1, 0)$. The size of the blue–yellow signal is surprisingly large given that only 7% of the cones are of the S type, and that neurons in the LGN that principally respond to S-cone stimulation are relatively rare^{16,17}. As the temporal frequency of the stimulus increases, the relative size of the responses to stimuli in the blue–yellow opponent direction and near the S-cone axis decreases compared to the responses to stimuli in the luminance directions.

We next compared the fMRI iso-response contours with psychophysical detection contours measured using the same checkerboard stimuli. The detection contours and individual threshold points for three temporal frequencies, two colour planes and one observer are shown in Fig. 4. Data from the other observer were in good agreement with these data, and both sets resemble previously published results^{3,14,15,18,19}. The solid contours represent threshold stimulus contrasts, and the dashed contours represent an 80% confidence interval.

The perceptual data are similar to the cortical data in several respects. At 1 and 4 Hz, in the red–green plane, sensitivity is highest to stimuli that cause opposing L- and M-cone signals. In the blue–yellow plane, relatively high sensitivity to S-cone stimuli falls rapidly as temporal frequency increases.

The main difference between the cortical and behavioural

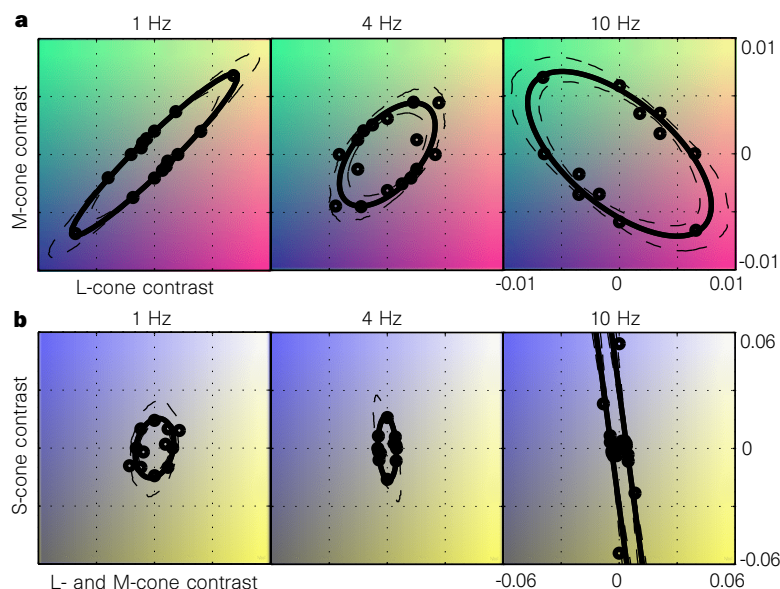


Figure 4 Psychophysical detection contours of the stimuli used in the fMRI experiments for observer B.W. The solid curve shows the detection threshold contour, and the open circles represent threshold contrast levels measured in various colour directions. The dotted curves represent a confidence interval, computed using a resampling procedure. **a**, Data in a red–green plane; **b**, data in a blue–yellow plane. Temporal frequencies at 1 Hz (left), 4 Hz (middle) and 10 Hz (right) are shown. Data for observer S.E. were similar.

measurements is found for stimuli at 10 Hz in the red–green plane. The strongest fMRI response is to stimuli that produce opposing L- and M-cone signals, but the behavioural sensitivity is best to stimuli in which the L- and M-cone signals covary. The fMRI data agree with preliminary electrophysiological data showing that some V1 neurons respond strongly to red–green stimuli at high temporal frequencies^{20,21}. Perceptual sensitivity to these stimuli may be limited by processing beyond areas V1 and V2.

These fMRI measurements are consistent with measurements of visual evoked potential (VEP) that found larger responses to chromatic stimuli than to luminance stimuli of comparable contrast^{22,23}. The cortical sources of the VEP signal are unknown, however. The present results are also in agreement with fMRI measurements showing a larger V1 response to a red–green chromatic stimulus than to a luminance stimulus²⁴. Our data extend these results by measuring complete colour-tuning curves, and comparing these results to behavioural data.

Our results indicate that cortical areas V1 and V2 are quite responsive to chromatic stimuli, and that the cortical responses resemble perceptual responses in important ways. These findings argue against the idea that perceptually relevant chromatic information is sparsely represented within V1 (refs 25, 26). The magnitude of the responses suggests that a relatively large population of neurons responds to the chromatic stimulation. The similarity between the cortical colour tuning and the behavioural colour sensitivity suggests that the cortical signals measured are relevant for colour perception. These kinds of colour-tuning measurements should allow us to learn how the spatiotemporal context, which influences colour perception in important ways, affects responses in the cortex. □

Methods

Identification of a visual areas. Our methods for identifying visual areas are reported elsewhere in detail^{27–30}; visual areas in the cortex were identified using measurements of retinotopic organization in the occipital lobe, viewed on flattened representations of cortex.

Scan planes. In all experiments, four or eight measurement planes of fMRI data were acquired using a blood oxygenation level-dependent pulse sequence (spiral *k*-space acquisition with echo time (TE) = 75, repetition time (TR) = 300, flip angle (FA) = 35, pixel size = $1.5 \times 1.5 \times 5$ mm. One image per plane was acquired every 2.5 s or every 3 s. Scan planes were oriented perpendicular to the calcarine sulcus and included the representation of the central 15 deg of the visual field. T1-weighted anatomical images were taken in register with each functional scan plane. Subjects' heads were immobilized using a bite-bar; no motion artefacts were detected in our data.

Stimuli. Stimuli were circular, 20 deg, contrast-reversing, checkerboard patterns surrounded by a neutral mean field. The stimuli were imaged using an LCD video projector. We defined our stimuli in terms of cone contrast. Formally, let the values L_0 , M_0 and S_0 represent the long-, medium- and short-wavelength cone absorptions from the mean field, and let ΔL , ΔM and ΔS , represent a perturbation of the stimulus checks from the mean value. The stimulus is represented by the three-dimensional cone contrast vector, $s = (\Delta L/L_0, \Delta M/M_0, \Delta S/S_0)$. Alternate checks were equal positive and negative perturbations. We define the colour contrast of the stimulus to be the vector length, $\|s\|$. The colour direction of the stimulus is unit length vector $s/\|s\|$.

Presentation of stimuli. Stimuli were presented in triplets with a common colour direction; the colour contrast between consecutive members of a triplet increased by a factor of two. Stimuli in each colour direction were presented twice, once in increasing order by contrast, and once in decreasing order. Six stimuli were presented during each fMRI scan.

fMRI contours. Iso-response contours were estimated using the average fMRI responses from all pixels in a cortical area. A criterion response was chosen that was approximately half of the maximum response measured in the experiment. Contours represent the interpolated set of stimuli that produce the criterion level of response. Confidence intervals were generated using a resampling procedure. We reanalysed our data 500 times, using random draws with replacement from our original data. Stimuli inside the outer contour produced responses greater than the criterion in 90% of the resamplings; stimuli inside the inner contour produced such responses in 10% of the resamplings.

Detection contours. Perceptual thresholds were measured using a display controlled by a 30-bit frame buffer. The probability of detecting each of many stimuli, which differed in colour direction and contrast, was measured using a two-interval, forced-choice, multiple staircase design. Thresholds for individual colour directions and the full elliptical detection contours were estimated using a maximum-likelihood procedure. The detection contours and confidence intervals represent estimates based on more than 1,000 trials. The inner and outer confidence contours represent 10% and 90% levels, based on the same resampling procedure used to analyse the fMRI data.

Received 17 January; accepted 30 April 1997.

- Hurvich, L. & Jameson, D. An opponent-process theory of color vision. *Psychol. Rev.* **64**, 384–404 (1957).
- Boynton, R. et al. Interactions among chromatic mechanisms as inferred from positive and negative increment thresholds. *Vis. Res.* **4**, 87–117 (1964).
- Thornton, J. & Pugh, E. Red/green color opponency at detection threshold. *Science* **219**, 191–193 (1983).
- Thorell, L. et al. Spatial mapping of monkey V1 cells with pure color and luminance stimuli. *Vis. Res.* **24**, 751–769 (1984).
- Livingstone, M. & Hubel, D. Anatomy and physiology of a colour system in the primate visual cortex. *J. Neurosci.* **4**, 2830–2835 (1984).
- Vautin, R. & Dow, B. Color cell groups in foveal striate cortex of the behaving macaque. *J. Neurophysiol.* **54**, 273–292 (1985).
- Lennie, P. et al. Chromatic mechanisms in striate cortex of macaque. *J. Neurosci.* **10**, 649–669 (1990).

8. DeValois, R. L. Analysis and coding of color vision in the primate visual system. *Cold Spring Harb. Symp. Quant. Biol.* **30**, 567–579 (1965).
9. Wiesel, T. & Hubel, D. S. Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey. *J. Neurophysiol.* **29**, 1115–1156 (1966).
10. Derrington, A. et al. Chromatic mechanisms in lateral geniculate nucleus of macaque. *J. Physiol. (Lond.)* **357**, 241–265 (1984).
11. Mullen, K. The contrast sensitivity of human colour vision to red–green and blue–yellow chromatic gratings. *J. Physiol. (Lond.)* **359**, 381–400 (1985).
12. Poirson, A. & Wandell, B. Appearance of colored patterns: Pattern-color separability. *J. Opt. Soc. Am.* **12**, 2458–2471 (1995).
13. Ts'o, C. & Gilbert, C. The organization of chromatic and spatial interactions in the primate striate cortex. *J. Neurosci.* **8**, 1712–1727 (1988).
14. Wandell, B. Color measurement and discrimination. *J. Opt. Soc. Am. A* **2**, 62–71 (1985).
15. Cole, G. et al. Detection of mechanisms in L-, M-, and S-cone contrast space. *J. Opt. Soc. Am. A* **10**, 38–51 (1993).
16. Curcio, C. A. et al. Distribution and morphology of human cone photoreceptors stained with anti-blue opsin. *J. Comp. Neurol.* **312**, 610–624 (1991).
17. Dacey, D. M. & Lee, B. B. The 'blue-on' opponent pathway in primate retina originates from a distinct bistratified ganglion cell type. *Nature* **367**, 731–735 (1994).
18. Kelly, D. Theory of flicker and transient responses: I. Uniform fields. *J. Opt. Soc. Am.* **61**, 537–546 (1971).
19. Stockman, A. The temporal properties of the human short-wave photoreceptors and their associated pathways. *Vis. Res.* **31**, 189–208 (1991).
20. Cottaris, N. P. et al. Spatio-temporal luminance and chromatic receptive field profiles of macaque striate cortex simple cells. *Soc. Neurosci. Abstr.* **22**, 951 (1996).
21. Gur, M. & Snodderly, D. M. A dissociation between brain activity and perception: Chromatically opponent cortical neurons signal chromatic flicker that is not perceived. *Vis. Res.* **37**, 377–382 (1997).
22. Regan, D. Evoked potentials specific to spatial patterns of luminance and colour. *Vis. Res.* **13**, 2381–2402 (1973).
23. Rabin, J. et al. Visual evoked potentials in three-dimensional color space: Correlates of spatio-chromatic processing. *Vis. Res.* **34**, 2657–2671 (1994).
24. Kleinschmidt, A. et al. Functional mapping of color processing by magnetic resonance imaging of responses to selective P- and M-pathway stimulation. *Exp. Brain Res.* **110**, 279–288 (1996).
25. Rodieck, R. W. in *From Pigments to Perception* (eds Valberg, A. & Lee, B. B.) 83–93 (Plenum, New York, 1991).
26. Calkins, D. J. & Sterling, P. Absence of spectrally specific lateral inputs to midget ganglion cells in primate retina. *Nature* **381**, 613–615 (1996).
27. Engel, S. A. et al. fMRI of human visual cortex. *Nature* **369**, 525 (1994).
28. Sereno, M. et al. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* **268**, 889–893 (1995).
29. DeYoe, E. et al. Mapping striate and extrastriate visual areas in human cerebral cortex. *Proc. Natl. Acad. Sci. USA* **93**, 2382–2386 (1996).
30. Engel, S. A. et al. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb. Cort.* **7**, 181–192 (1997).

Acknowledgements. We thank G. Glover for technical help, and H. Baseler, G. Boynton, E. J. Chichilnisky, E. Markman, W. Newsome, A. Poirson and J. Wine for comments on the manuscript. This work was funded by the National Eye Institute and the McDonnell-Pew Foundation.

Correspondence and requests for materials should be addressed to S.A.E. (e-mail: engel@psych.ucla.edu).

Hippocampal GABA_A channel conductance increased by diazepam

M. Eghbali, J. P. Curmi, B. Birnir & P. W. Gage

John Curtin School of Medical Research, PO Box 334, Australian National University, Canberra, ACT 2601, Australia

Benzodiazepines, which are widely used clinically for relief of anxiety and for sedation¹, are thought to enhance synaptic inhibition in the central nervous system by increasing the open probability of chloride channels activated by the inhibitory neurotransmitter γ -aminobutyric acid (GABA)^{2,3}. Here we show that the benzodiazepine diazepam can also increase the conductance of GABA_A channels activated by low concentrations of GABA (0.5 or 5 μ M) in rat cultured hippocampal neurons. Before exposure to diazepam, chloride channels activated by GABA had conductances of 8 to 53 pS. Diazepam caused a concentration-dependent and reversible increase in the conductance of these channels towards a maximum conductance of 70–80 pS and the effect was as great as 7-fold in channels of lowest initial conductance. Increasing the conductance of GABA_A channels tonically activated by low ambient concentrations of GABA in the extracellular environment⁴ may be an important way in which these drugs depress excitation in the central nervous system. That any drug has such a large effect on single channel conductance has not been reported previously and has implications for models of channel structure and conductance.

Cultures of hippocampal neurons from neonatal rats were prepared as described previously⁵ and single-channel currents activated by 0.5 or 5 μ M GABA were recorded in cell-attached, inside-out or outside-out patches. Currents recorded had properties typical of GABA-activated chloride channels in these cells^{6–8}: they reversed at a pipette potential close to 0 mV even when Na⁺ was replaced by K⁺ in the pipette solution, showed outward rectification, had conductances that varied over a wide range and were blocked by injection of 100 μ M bicuculline into the pipette tip. In 27 patches in which single-channel currents were recorded before and after exposure to diazepam, single-channel conductance (current amplitude divided by the difference between the pipette potential (–40 to –80 mV) and the reversal potential) before exposure of a patch to diazepam ranged from 8 to 53 pS (Table 1). This is a wide variation but within the range of conductances of GABA_A channels^{9,10}.

The binding site for benzodiazepines is thought to involve residues in the extracellular domain of α and perhaps γ -subunits¹¹. When diazepam was applied to the extracellular surface of cell-attached patches by injecting the drug into the pipette tip⁶, an increase in the amplitude of single-channel currents activated by GABA was seen. In Fig. 1, single channel currents activated by 5 μ M GABA (Fig. 1a) had an average amplitude of 0.9 pA (pipette potential of –80 mV), as can be seen in the all-points histogram in Fig. 1c. Following injection of a bolus of control solution containing 5 μ M GABA plus 10 μ M diazepam into the pipette tip, the single-channel currents (Fig. 1b) showed more than a fourfold increase in amplitude, giving a single open peak in the all-points

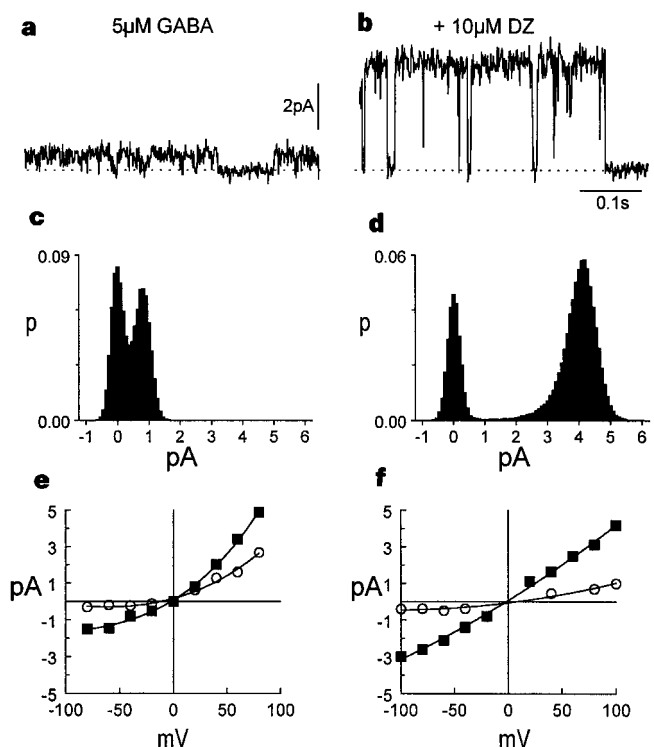


Figure 1 The increase in single-channel current amplitude caused by 10 μ M diazepam in a cell-attached patch (pipette potential of –80 mV). **a**, Single-channel currents activated by 5 μ M GABA. **b**, Single-channel currents recorded from the same patch after injection of a solution containing 5 μ M GABA and 10 μ M diazepam into the pipette tip. **c**, All-points histogram of a 16-s current recording, including the segment in **a**. **d**, All-points histogram of a 16-s current recording including the segment in **b**. **e**, Current–voltage relation for single-channel currents activated by 0.5 μ M GABA in a cell-attached patch before (circles) and after (squares) injection of 1 μ M diazepam. **f**, Current–voltage relation for single-channel currents activated by 0.5 μ M GABA in an inside-out patch before (circles) and after (squares) exposure to 1 μ M diazepam.

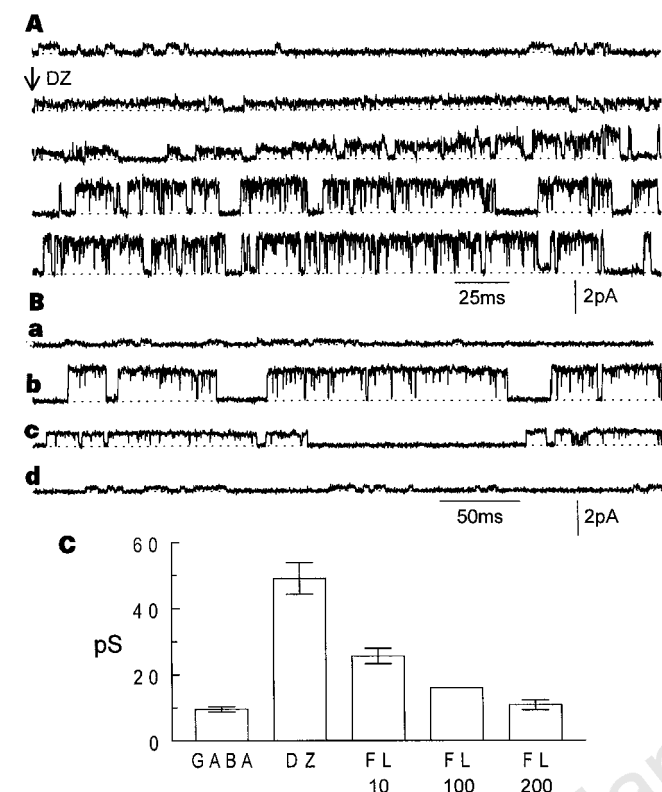


Figure 2 **A** A gradual increase in single-channel current amplitude caused by diazepam in an inside-out patch (pipette potential of -80 mV). Top trace shows currents activated by $5 \mu\text{M}$ GABA alone. At the arrow at the beginning of the second trace, bath solution containing GABA plus $1 \mu\text{M}$ diazepam began to flow into the bath. The 4 traces after the arrow are part of a continuous record. **B**, Flumazenil reverses the increase in channel conductance caused by diazepam in an inside-out patch (pipette potential of -40 mV). **a**, $0.5 \mu\text{M}$ GABA; **b**, $0.5 \mu\text{M}$ GABA plus $1 \mu\text{M}$ diazepam; **c**, $0.5 \mu\text{M}$ GABA plus $1 \mu\text{M}$ diazepam plus 10 nM flumazenil; **d**, $0.5 \mu\text{M}$ GABA plus $1 \mu\text{M}$ diazepam plus 200 nM flumazenil. **C**, Conductance of channels (mean $\pm$ 1 s.e.m.) activated in 7 inside-out patches by $0.5 \mu\text{M}$ GABA (GABA) was first increased by $1 \mu\text{M}$ diazepam (DZ) and then depressed by $10 \mu\text{M}$ flumazenil (FL 10). One of the patches was then exposed to 100 nM (FL 100), and 4 to 200 nM (FL 200) flumazenil.

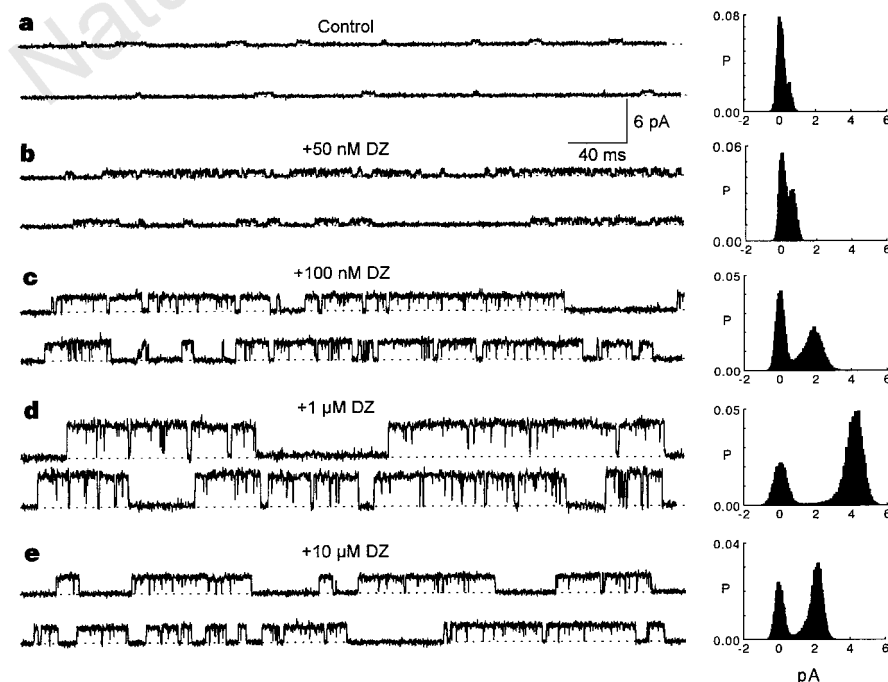


Figure 3 The influence of diazepam concentration on single-channel currents activated by $0.5 \mu\text{M}$ GABA. All currents were recorded from the same inside-out patch during exposure of the patch to successively higher concentrations of diazepam from 50 nM (**a**) to $10 \mu\text{M}$ (**e**), delivered to the patch through the bath. For all records, the calibrations between **a** and **b** apply and the pipette potential was -60 mV. Current amplitude probability histograms shown alongside traces **a-e** were obtained from 16-s records digitized at 10 kHz .

histogram (Fig. 1d) close to 4 pA . The effect was not due to the injection technique nor to the dimethylsulphoxide (DMSO) used: injection into the pipette tip of pipette solution of GABA plus $190 \mu\text{M}$ DMSO caused no increase in single-channel current amplitude ($n = 6$). Nor was the increase in current amplitude caused by a change in reversal potential (Fig. 1e). Therefore the diazepam caused a fourfold increase in single-channel conductance (Fig. 1a–d). The increase in single-channel conductance (γ) caused by concentrations of diazepam from 100 nM to $10 \mu\text{M}$ in 16 cell-attached patches can be seen in Table 1 (single asterisks). Also shown in Table 1 is the mean conductance $\bar{\gamma}$ (average current over a 16-s period divided by (pipette potential – reversal potential)), which reflects both channel conductance and open probability. In most patches, diazepam caused an increase in $\bar{\gamma}$ greater than could be explained by the increase in channel conductance, and this must have been due to an increase in channel-open probability, as reported previously^{2,3}. An increase in the number of channels active in a patch was sometimes apparent from the appearance of superimposed single-channel currents and the numbers of channels detected in this way before and after exposure to diazepam are also given in Table 1. It can be concluded that diazepam was causing an increase in both single-channel conductance and channel open probability. The relative contributions of these two factors to the increase in mean chloride conductance caused by diazepam varied from patch to patch, as can be seen in the last two columns of Table 1.

Diazepam is lipid-soluble and should pass through cell membranes. In 5 experiments on cell-attached patches, an increase in the amplitude of single-channel currents occurred several minutes after diazepam (0.1 to $10 \mu\text{M}$) was added to the bath solution (not shown). It was assumed that the diazepam had reached its site of action by diffusing into the cell and back across the membrane patch. To avoid the injection technique, inside-out patches were exposed to diazepam in the flowing bath solution. Slowly running the diazepam solution into the bath gave a progressive increase in channel conductance (see example in Fig. 2). At the arrow at the beginning of the second trace, bath solution containing $1 \mu\text{M}$ diazepam began to flow into the bath. A gradual increase in the amplitude of single-channel currents is apparent in the 4 continuous traces after switching to the diazepam solution (Fig. 2). Again, diazepam did not change the reversal potential of currents recorded in inside-out patches (Fig. 1f), but increased channel conductance.

The effects of diazepam could be reversed by washing the drug out of the bath (not shown).

The increase in channel conductance caused by diazepam could be reversed with the benzodiazepine antagonist flumazenil (RO 15-1788, a gift from Hoffmann-La Roche) (Fig. 2B,C). The conductance of channels activated by 0.5 μ M GABA increased from 7 pS (Fig. 2B,a) to 52 pS (Fig. 2B,b) when the patch was exposed to 1 μ M diazepam. Upon exposure of the patch to the diazepam together with 10 nM flumazenil, channel conductance decreased to 25 pS (Fig. 2B,c). Raising the flumazenil concentration to 200 nM further depressed channel conductance to 7 pS (Fig. 2B,d). Similar results obtained in 7 experiments are shown in Fig. 2C.

The progressive increase in channel conductance during slow bath-application of diazepam and the effect of the antagonist indicated that the conductance of the channel might be related to diazepam concentration. The relationship between diazepam concentration and channel conductance was therefore examined more closely. Results obtained in an experiment in which diazepam concentration in the bath was varied from 50 nM to 10 μ M are shown in Fig. 3. The single-channel currents recorded from the inside-out patch (pipette potential -60 mV) before exposure to diazepam had an average amplitude of 0.65 pA (10.8 pS). Following exposure to 50 nM diazepam applied to the inside surface of the membrane in the bath solution (Fig. 3b), currents increased in amplitude to 1.32 pA (22 pS). Single-channel current amplitude increased to 2.34 pA (39 pS) when diazepam concentration was increased to 100 nM (Fig. 3c) and to 4.76 pA (79 pS) at a diazepam concentration of 1 μ M (Fig. 3d). However, increasing diazepam concentration further to 10 μ M (Fig. 3e) gave smaller single-channel currents (2.1 pA, 35 pS) than with 1 μ M diazepam.

Similar results were obtained in six inside-out patches by varying diazepam concentration. The average ratio of single-channel conductance in the presence of diazepam to that recorded before

exposure of the patch to diazepam ($\gamma(\text{DZ})/\gamma(\text{Con})$) is plotted against diazepam concentration for the 6 patches in Fig. 4a. There was no detectable change in channel conductance with 10 or 20 nM diazepam, but a progressive increase as diazepam concentration increased from 50 nM to 1 μ M. When diazepam concentration was increased to 10 μ M, channel conductance was depressed. These results and the effects of diazepam in two outside-out patches are included in Table 1.

The fact that diazepam can cause a large increase in the conductance of GABA_A channels has not been reported previously^{3,11,13}. There are several possible reasons for this. Firstly, our experiments were on hippocampal neurons. The effect of diazepam is known to be subunit-dependent¹¹ and the receptors we studied may have a subunit composition and response to diazepam different from receptors examined by others. Second, diazepam had only a small effect on channels that had a conductance above ~20 pS. This can be seen in Table 1 but is clearer in Fig. 4b in which the relative increase in channel conductance ($\gamma(\text{DZ})/\gamma(\text{Con})$) is plotted against channel conductance before exposure to diazepam ($\gamma(\text{Con})$). There is an inverse relationship between the increase in channel conductance and the initial channel conductance. If the initial conductance was below 20 pS, diazepam always caused a large increase in channel conductance but the effect was less in channels with higher initial conductance. Indeed, channels activated with 100 μ M GABA had a high conductance of 64.4 ± 5.72 pS (mean $\pm$ 1 s.e.m., $n = 8$) and there was no change in channel conductance when two of these patches were exposed to 1 μ M diazepam. Third, the conditions of our experiments were different from others in some respects: for example, we always had ATP at the inner membrane surface. The state of the receptor (for example, whether it is phosphorylated) might modify the effect of diazepam.

Just as the mechanism by which benzodiazepines alter channel kinetics is not known, we do not know how diazepam increases

Table 1 Effects of diazepam on GABA_A channels

Patch	[GABA] (μ M)	γ (pS)	$\bar{\gamma}$ (pS)	No. of channels	+DZ (μ M)	γ (pS)	$\bar{\gamma}$ (pS)	No. of channels	Ratio γ	Ratio $\bar{\gamma}$
1*	5	11.1	5.13	1	10	56.3	40.4	1	5.06	7.88
2*	5	36.3	29.3	1	10	58.8	277	6	1.62	9.46
3*	5	22.5	11.0	1	10	60.0	37.9	1	2.67	3.44
4*	5	38.8	22.1	1	10	61.3	32.4	1	1.58	1.46
5*	5	35.0	18.5	1	10	61.3	84.6	2	1.75	4.57
6	5	11.8	2.88	1	10	35.0	9.25	2	2.98	3.22
7a	5	8.33	3.83	1	10	34.7	16.0	1	4.16	4.17
8*	0.5	11.3	3.50	1	10	62.5	21.0	1	5.56	6.00
9*	0.5	18.8	1.38	1	10	52.5	32.1	1	2.80	23.4
10*	0.5	50.0	36.6	1	10	70.0	50.3	1	1.40	1.37
11a	0.5	10.8	0.33	1	10	34.7	24.0	1	3.20	72.7
12a	0.5	16.7	3.00	1	10	41.6	10.1	1	2.49	3.37
7b	5	8.33	3.83	1	1	54.5	27.2	1	6.54	7.09
13**	5	24.0	2.23	1	1	46.0	21.3	1	1.92	9.55
14**	1	17.0	1.89	1	1	64.0	31.2	1	3.76	16.5
15*	0.5	43.8	23.3	1	1	61.3	102	3	1.40	4.39
16*	0.5	40.0	30.6	1	1	60.0	49.5	1	1.50	1.62
17*	0.5	52.5	37.8	1	1	66.3	74.4	2	1.26	1.97
18*	0.5	33.8	19.6	1	1	61.3	50.3	2	1.81	2.56
19*	0.5	29.5	16.3	1	1	80.0	27.0	1	2.71	1.65
11b	0.5	10.8	0.33	1	1	79.3	48.8	1	7.33	147
20a	0.5	17.0	0.83	1	1	64.8	9.85	1	3.81	11.9
12b	0.5	16.7	3.00	1	1	70.0	25.8	1	4.19	8.61
21	5	19.8	9.87	1	0.1	35.8	19.8	1	1.81	2.00
22a	5	12.3	1.58	1	0.1	51.3	7.00	1	4.16	4.42
23	5	17.0	3.45	1	0.1	53.0	12.1	1	3.12	3.51
24*	0.5	16.3	9.50	1	0.1	73.8	257	6	4.54	27.1
25*	0.5	23.8	13.4	1	0.1	41.3	18.9	1	1.74	1.41
11c	0.5	10.8	0.33	1	0.1	39.0	16.8	1	3.60	50.5
26a	0.5	13.7	2.00	1	0.1	34.3	76.3	3	2.51	38.2
20b	0.5	17.0	0.83	1	0.1	40.2	3.53	1	2.36	4.25
22b	5	12.3	1.58	1	0.05	27.7	6.17	1	2.24	3.89
27	5	44.0	8.56	1	0.05	51.3	10.9	1	1.16	1.27
11d	0.5	10.8	0.33	1	0.05	22.0	6.17	2	2.03	18.5
26b	0.5	13.7	2.0	1	0.05	24.8	35.8	2	1.82	17.9

Single-channel conductance (γ , single-channel current/(pipette potential - reversal potential)), mean conductance ($\bar{\gamma}$, mean 16-s current/(pipette potential - reversal potential)) and the number of channels opening in 15 cell-attached (one asterisk), 10 inside-out patches and 2 outside-out patches (two asterisks) activated by 0.5 or 5 μ M GABA, before and after exposure to diazepam at 0.05–10 μ M. Results from the same patch are given suffix a, b or c. The post-drug/pre-drug ratios of single channel or mean conductances are shown in the last two columns. Pipette potentials were -40, -60 or -80 mV.

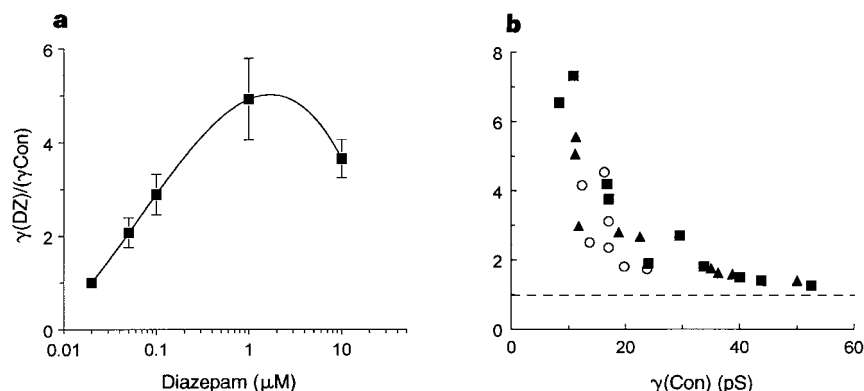


Figure 4 **a**, The relation between diazepam concentration and average single-channel conductance in 5 inside-out patches. Conductance is normalized to the conductance recorded before exposure to diazepam ($(\gamma(\text{DZ})/\gamma(\text{Con}))$; ordinate). The curve is a best fit third-order polynomial. Vertical bars show ± 1 s.e.m. **b**, The relation between control (pre-drug) single channel conductance ($\gamma(\text{Con})$) and the maximum relative conductance (ordinate as in **a**) caused by diazepam in each of the 27 patches. Diazepam concentrations are: triangles, 0.1 μM ; squares, 1 μM ; circles, 10 μM .

channel conductance. It may induce a different, higher-conductance, open-channel conformation of the channel protein, or it may produce synchronous opening of a number of channels of low conductance. There is good evidence that a chloride channel (ClC-0)^{14–16} and an inward rectifier K channel^{17,18} are multibarrelled structures (multi-pores). Because of the equal voltage-independent spacing of subconductance levels in GABA_A channels, it has been suggested⁸ that they also may be multibarrelled with variable voltage-dependent coupling of the barrels (pores), giving outward rectification. The graded effect of diazepam (Fig. 4a) would require multiple binding sites for diazepam, either on one receptor or on many receptors. Unfortunately, the stoichiometry of benzodiazepine binding to a GABA_A receptor is not known¹¹. We favour the idea (which is by no means proven) that the multiple diazepam binding sites are on multiple receptors and that diazepam increases conductance by promoting synchronous opening and closing of clustered pores.

Benzodiazepines have only a small effect on the peak amplitude of inhibitory postsynaptic currents (IPSCs) in the hippocampus^{19–21} and this is consistent with the idea that the conductance of channels activated by high concentrations of GABA would not be increased by diazepam. However, the prolonged decay of IPSCs caused by benzodiazepines^{19–21} could be due to maintenance of high-conductance channels as GABA dissociates from receptors. The major effect of benzodiazepines may not be on synaptic transmission but on tonic inhibition⁴ mediated by extrasynaptic GABA_A receptors. There is evidence that the extracellular concentration of GABA in the central nervous system is 0.2 to 0.8 μM ^{22,23} and it has been proposed⁴, from observations on rat dentate gyrus neurons, that this would produce a tonic background of extrasynaptic GABA_A channel activity that would oppose excitation. Our results here indicate that the conductance of such channels in the hippocampus tonically activated by ambient concentrations of GABA would be increased significantly by diazepam. □

Methods

Hippocampal neurons were obtained from neonatal rats and maintained in culture for 5 to 12 days⁵. The 'bath' solution and pipette solution normally contained (in mM): NaCl 135, KCl 3, CaCl₂ 2, MgCl₂ 5, N-tris(hydroxymethyl)methyl-2-amino ethane sulphonic acid (TES) 10. ATP (4 mM) was present in solutions to which the inner membrane surface was exposed. The pH of solutions was adjusted to 7.3 with NaOH. In some experiments, the pipette solution contained KCl instead of NaCl. Diazepam and flumazenil were dissolved in DMSO then in bath solution to give a stock solution of 100 μM diazepam or flumazenil in 19 mM DMSO. Aliquots of this stock solution were added to pipette or bath solutions to give the required concentrations. Cells on glass coverslips were transferred for experiments to a plastic bath on the stage of an inverted microscope. The bath solution flowed constantly through the bath from one end to the other, where it was removed by suction. Experiments were done at room temperature (20–22 °C). Pipettes were made from borosilicate

glass (Clark Electromedical) coated with Sylgard (Dow Corning) and fire-polished. They had a resistance of 5 to 15 M Ω when filled with bath solution.

Currents were recorded with an Axopatch 1D amplifier (Axon Instruments), filtered at 2 or 5 kHz, digitized at 44 kHz (Sony PCM) and stored on video tape. For analysis, currents were played back using the same system and digitized at different frequencies, 10 kHz or lower, using a Tecmar A to D converter interfaced with an IBM-compatible PC. For analysis of single-channel currents, records were played back from the VCR onto an oscilloscope and chart recorder for scanning. Representative segments were then selected and digitized, normally at 10 kHz, to obtain a 320 kB file. These segments were then displayed using an IBM PC-compatible computer, the closed state set at zero by visual inspection of the whole file and a current amplitude probability histogram of the data points constructed, normally with a bin width of 0.1 pA (program written by M. Smith). Outward currents across the membrane (inward chloride movement) are shown as positive, upward currents and inward currents are shown as negative, downward currents. The amplitude of single-channel currents was measured from peaks in all-points histograms and by direct measurement of single-channel current amplitude.

Received 10 March; accepted 29 April 1997.

1. Tanelian, D. L., Kosek, P., Mody, I. & MacIver, M. B. The role of the GABA_A receptor/chloride channel complex in anesthesia. *Anesthesiology* **78**, 757–776 (1993).
2. Study, R. E. & Barker, J. L. Diazepam and (–)-pentobarbital: fluctuation analysis reveals different mechanisms for potentiation of γ -aminobutyric acid responses in cultured central neurons. *Proc. Natl Acad. Sci. USA* **78**, 7180–7184 (1981).
3. Rogers, C. J., Twyman, R. E. & MacDonald, R. L. Benzodiazepine and β -carboline regulation of single GABA_A receptor channels of mouse spinal neurones in culture. *J. Physiol. (Lond.)* **475**, 69–82 (1994).
4. Birnir, B., Everitt, A. B. & Gage, P. W. Characteristics of GABA_A channels in rat dentate gyrus. *J. Membr. Biol.* **142**, 93–102 (1994).
5. Premkumar, L. S., Chung, S.-H. & Gage, P. W. GABA-induced potassium channels in cultured neurons. *Proc. R. Soc. Lond. B* **241**, 153–158 (1990).
6. Curmi, J. P., Premkumar, L. S., Birnir, B. & Gage, P. W. The influence of membrane potential on chloride channels activated by GABA in rat cultured hippocampal neurones. *J. Membr. Biol.* **136**, 2673–280 (1993).
7. Fatima-Shad, K. & Barry, P. H. A patch-clamp study of GABA- and strychnine-sensitive glycine-activated currents in post-natal tissue-cultured hippocampal neurons. *Proc. R. Soc. Lond. B* **250**, 99–105 (1992).
8. Gage, P. W. & Chung, S. Influence of membrane potential on conductance sublevels of chloride channels activated by GABA. *Proc. R. Soc. Lond. B* **255**, 167–172 (1994).
9. Gray, R. & Johnston, D. Rectification of single GABA-gated chloride channels in adult hippocampal neurons. *J. Neurophysiol.* **54**, 134–142 (1985).
10. Mathers, D. A. Activation and inactivation of the GABA_A receptor: insights from comparison of native and recombinant subunit assemblies. *Can. J. Physiol. Pharmacol.* **69**, 1057–1063 (1991).
11. Smith, G. B. & Olsen, R. W. Functional domains of GABA_A receptors. *Trends Pharmacol. Sci.* **16**, 162–168 (1995).
12. Vicini, S., Mienville, J. M. & Costa, E. Actions of benzodiazepine and β -carboline derivatives on γ -aminobutyric acid-activated Cl[–] channels recorded from membrane patches of neonatal rat cortical neurons in culture. *J. Pharmacol. Exp. Ther.* **243**, 1195–1201 (1987).
13. Twyman, R. E., Rogers, C. J. & MacDonald, R. L. Differential regulation of γ -aminobutyric acid receptor channels by diazepam and phenobarbital. *Ann. Neurol.* **25**, 213–220 (1989).
14. Miller, C. Open-state substructure of single chloride channels from *Torpedo* electroplax. *Phil. Trans. R. Soc., Lond. B* **299**, 401–411 (1982).
15. Middleton, R. E., Pheasant, D. J. & Miller, C. Homodimeric architecture of a ClC-type chloride ion channel. *Nature* **383**, 337–340 (1996).
16. Ludewig, W., Pusch, M. & Jentsch, T. J. Two physically distinct pores in the dimeric ClC-0 chloride channel. *Nature* **383**, 340–343 (1996).
17. Matsuda, H. Open-state substructure of inwardly rectifying potassium channels revealed by magnesium block in guinea-pig heart cells. *J. Physiol. (Lond.)* **397**, 237–258 (1988).
18. Matsuda, H., Matsuura, H. & Noma, A. Triple-barrel structure of inwardly rectifying K⁺ channels revealed by Cs⁺ and Rb⁺ block in guinea-pig heart cells. *J. Physiol. (Lond.)* **413**, 139–157 (1989).
19. Segal, M. & Barker, J. L. Rat hippocampal neurons in culture: voltage-clamp analysis of inhibitory synaptic connections. *J. Neurophysiol.* **52**, 469–487 (1984).
20. Otis, T. S. & Mody, I. Modulation of decay kinetics and frequency of GABA_A receptor-mediated

- spontaneous inhibitory postsynaptic currents in hippocampal neurons. *Proc. Natl Acad. Sci. USA* **78**, 7180–7184 (1992).
21. Mody, I., De Doninck, Y., Otis, T. S. & Soltesz, I. Bridging the cleft at GABA synapses in the brain. *Trends Neurosci.* **17**, 517–525 (1994).
22. Tossman, U., Jonsson, G. & Understedt, U. Regional distribution and extracellular levels of amino acids in rat central nervous system. *Acta Physiol. Scand.* **127**, 533–545 (1986).
23. Lerma, J., Herranz, A. S., Herreras, O., Abraira, V. & Martin del rio, R. *In vivo* determination of extracellular concentration of amino acids in the rat hippocampus. A method based on brain dialysis and computerized analysis. *Brain Res.* **384**, 145–155 (1986).

Acknowledgements. We thank Graeme Cox for discussion and helpful comments.

Correspondence and requests for materials should be addressed to P.W.G.

p47 is a cofactor for p97-mediated membrane fusion

Hisao Kondo*, Catherine Rabouille*, Richard Newman†, Timothy P. Levine‡, Darryl Pappin§, Paul Freemont† & Graham Warren*

* Cell Biology Laboratory, † Protein Structure Laboratory, and § Protein Sequencing Laboratory, Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK

‡ MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

At least two distinct ATPases, NSF and p97, are known to be involved in the heterotypic fusion of transport vesicles with their target membranes and the homotypic fusion of membrane compartments¹. The NSF-mediated fusion pathway is the best characterized, many of the components having been identified and their functions analysed^{2–7}. In contrast, none of the accessory proteins for the p97-mediated fusion pathway has been identified^{8–10}. Now we have identified the first such component, a protein of relative molecular mass 47,000 (p47), which forms a tight, stoichiometric complex with cytosolic p97 (one trimer of p47 per hexamer of p97). It is essential for the p97-mediated regrowth of Golgi cisternae from mitotic Golgi fragments, a process restricted to animal cells¹¹. As a homologue of p47 exists in budding yeast, this indicates that it might also be involved in other membrane fusion reactions catalysed by p97, such as karyogamy¹⁰.

p97 is a hexameric protein that was originally purified on the basis of its abundance and large size. It comprises about 1% of the cell cytosol and has a sedimentation coefficient of 14.5S (refs 12, 13). The sequence shows extensive homology with NSF¹³ but it catalyses a different set of fusion reactions. p97 is involved in the fusion of outer nuclear envelopes during yeast karyogamy¹⁰ and in the reassembly of Golgi cisternae from fragments generated by treatment with specific drugs⁹ or mitotic cytosol⁸. p97 may also be involved in cell division because there is a homologue in archaeobacteria (*Sulfolobus acidocaldarius*)¹⁴ (*Halobacterium salinarum*; GenBank accession no. X79560) and these organisms have no internal membranes. The only cytoplasmic fusion event they need occurs during cytokinesis and leads to cell separation. There is no obvious homologue of NSF in archaeobacteria, consistent with its role in the internal membrane traffic events of higher eukaryotes. This also suggests that p97 is the more ancient of the two fusion ATPases.

We observed a consistent difference in the relative molecular mass (M_r) of p97 when the pure protein from rat liver was compared with that in crude cytosol using sucrose gradients (data not shown) and gel filtration columns (Fig. 1). On Superose-6 columns the pure protein had an apparent M_r of 600 ± 20 K (s.e.m.; $n = 3$), very similar to that reported for the *Xenopus* protein (570 ± 10 K; ref. 13). The p97 in crude rat liver cytosol was 140K larger, with an apparent M_r of 740 ± 30 K (s.e.m.; $n = 3$) (Fig. 1a). As the entire peak was shifted, this argued that all the p97 in crude cytosol was

complexed to a protein or proteins that were removed during subsequent purification.

We had previously noticed a protein of 47K (p47) that was difficult to remove completely during the purification of p97 (see Fig. 2a, lane 1 for example)⁸. Enough of this protein was extracted from gel slices to obtain partial sequences by mass spectrometry; peptide antibodies were raised and affinity-purified. When used to probe the fractionated cytosolic proteins, all of the p47 cofractionated with cytosolic p97 (compare Fig. 1a,b).

An association between p47 and p97 in crude cytosol was confirmed by immunoprecipitation using anti-p47 peptide antibodies. These antibodies immunoprecipitated p97, as revealed by western blotting (Fig. 2b, lane 2) but not when the antibodies were preblocked with the peptide (lane 3) or when the preimmune serum was used (lane 1).

Separation of p47 from purified p97 (Fig. 2a, lane 1) using gel filtration was achieved by raising the salt concentration to 1M. This yielded pure p97 (Fig. 2a, lane 2) and pure p47, which had an apparent M_r of 140 ± 30 K (s.e.m.; $n = 3$) by gel filtration (Fig. 1b). As p47 was unstable and yields were low, we prepared a recombinant p47. Peptide 7, the longest p47 peptide, was used as the starting point for a full-length complementary DNA (see Methods); the predicted open reading frame is shown in Fig. 3a.

No upstream stop codon was found so *in vitro* transcription and translation were used to show that initiation from the putative start methionine gave a protein of the same M_r as native p47 (data not shown). The full-length cDNA was modified to include a His₆ tag at the N terminus and the tagged p47 was expressed and purified on nickel columns (Fig. 2a, lane 3). Recombinant His₆-p47 still bound p97, as shown by adding the protein to crude cytosol followed by antibodies against the His₆ tag. The immunoprecipitate was then fractionated by SDS-PAGE and the gel stained with Coomassie blue. As shown in Fig. 2c, p97 appeared to be the only cytosolic protein that bound to His₆-p47 (compare lanes 1 and 2). This argues that p97 is the only cytosolic binding partner for p47 and that endogenous p47 can be exchanged for exogenous p47.

Recombinant His₆-p47 had the same apparent M_r as the native protein, as determined by gel filtration (140K; data not shown) and this was confirmed by light-scattering experiments (M_r 148 ± 13 K; $n = 4$). These results, together with those from crosslinking studies (not shown), indicate that p47 is a trimer. As the size of p47 is similar to the difference in apparent M_r between pure and cytosolic p97 by gel filtration (140K; Fig. 1a), this suggests that a trimer of p47

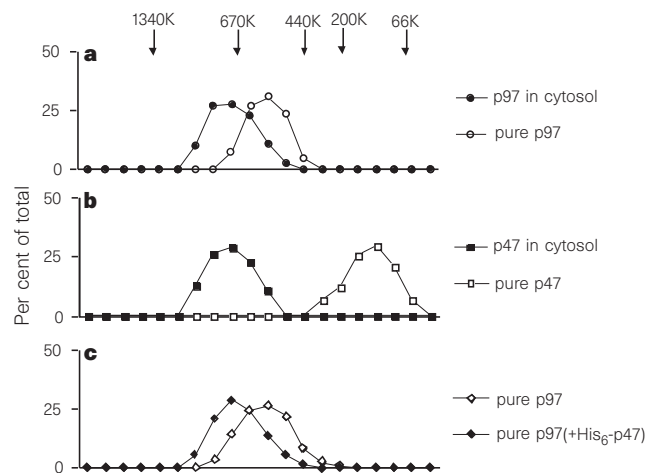


Figure 1 Gel filtration of p47 and p97. The indicated samples were fractionated on Superose 6 columns and fractions assayed by Western blotting and quantitative ECL. The average of three runs is shown for each sample in **a** and **b** and two runs in **c**.

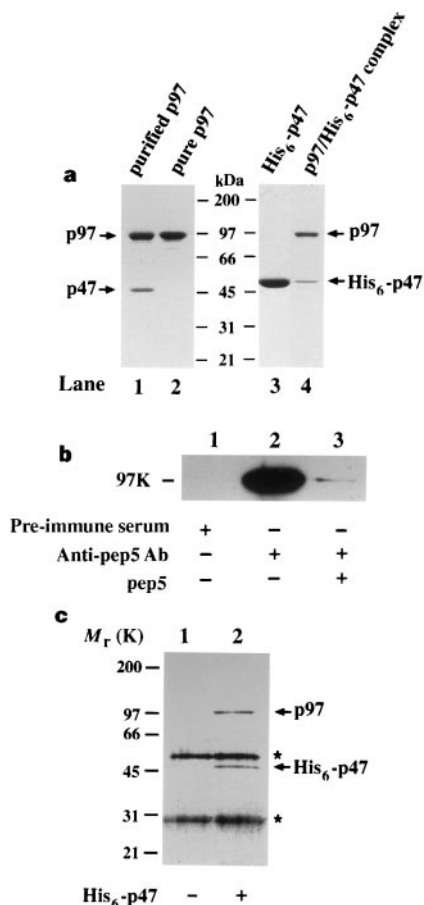


Figure 2 Association of p47 with p97. **a**, Lanes: 1, purified p97 (1.5 µg); 2, pure p97 (1.5 µg); 3, His₆-p47 (2.0 µg); 4, incubation of pure p97 (10 µg) and His₆-p47 (30 µg) followed by isolation of the complex by gel filtration (Fig. 1c). After SDS-PAGE, gels were stained with Coomassie blue. **b**, Rat liver cytosol (1 mg) was incubated with preimmune serum or antibodies against peptide 5 (blocked in one experiment with peptide 5). The immunoprecipitates were fractionated by SDS-PAGE followed by western blotting using anti-p97 antibodies. **c**, Rat liver cytosol (0.75 mg) was incubated in the presence or absence of His₆-p47 (3 µg) followed by immunoprecipitation with anti-His antibodies. After SDS-PAGE, the gel was stained with Coomassie blue. The asterisk refers to the light and heavy antibody chains.

binds to a hexamer of p97. We confirmed this by light-scattering measurements which showed that the difference in M_r between pure p97 ($590 \pm 30K$; $n = 2$) and reconstituted p97/p47 complex ($730 \pm 40K$; $n = 2$) was also 140K. This reconstituted complex was prepared by mixing pure p97 with excess His₆-p47 and isolated by gel filtration (Fig. 1c). The apparent M_r of 720K ($n = 2$) was very similar to that obtained using light scattering and also to that of cytosolic p97 (compare Fig. 1a,c), arguing that p47 is the only major protein complexed to p97 in rat liver cytosol.

The molar ratio of His₆-p47 to p97 was also calculated using the Coomassie stained bands in the peak fraction of the p97/p47 complex (Fig. 2a, lane 4). The result was 1 mol p47 to 2.0 mol p97; the molar ratio was the same when recombinant His₆-p47 was incubated with excess p97 and the complex isolated using anti-His antibodies (data not shown).

The structure of the reconstituted complex was examined by negative staining. The six subunits of pure p97 make up the staves of a barrel-shaped structure which is 12 nm wide, with a central channel of ~ 3.5 nm in diameter (Fig. 4a)¹². After reconstitution

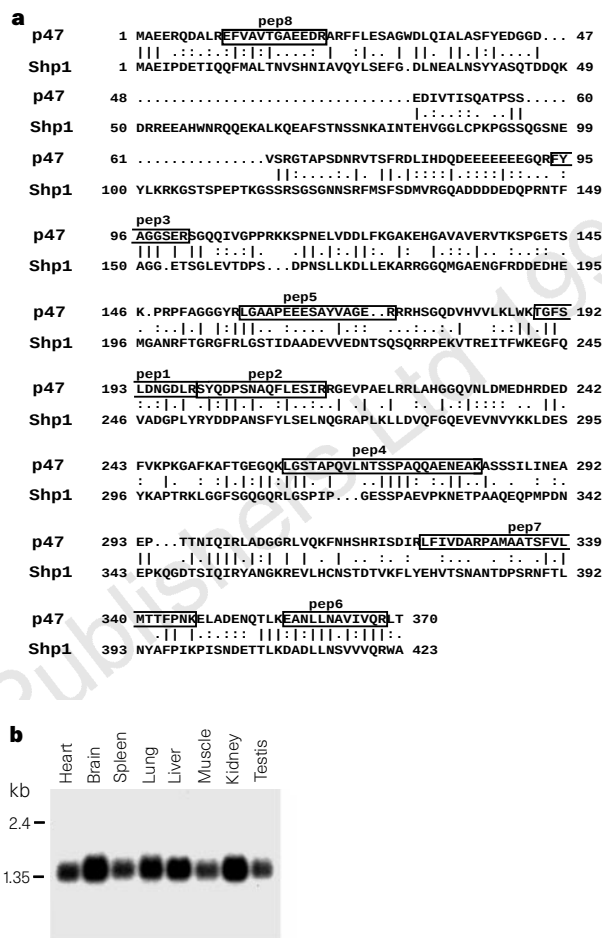


Figure 3 Predicted sequence of p47 and the yeast homologue Shp1. **a**, The predicted open reading frame of the full-length rat cDNA compared with the yeast homologue Shp1. The eight peptide sequences derived from the gel-purified p47 are boxed. **b**, Northern blot using full-length cDNA to probe a rat tissue poly(A)⁺ library.

with His₆-p47, the diameter of the central channel was reduced to ~ 1.2 nm (Fig. 4b), consistent with a p47 trimer being located at one end of the p97 barrel.

Northern blotting showed that p47 was present in all rat tissues tested, revealing a single transcript of 1.5 kilobases (kb) (Fig. 3b). More than 88% of the sequence is covered by the human expressed sequence tag (EST) database (IMAGE Consortium/U. Washington EST project) and the sequence identity is $\sim 89\%$. A homologue was found in the budding yeast *Saccharomyces cerevisiae* that has 52% similarity and 30% identity extending over the whole protein (Fig. 3a). Disruption of this homologue, termed Shp1, slows growth considerably¹⁵. Mutations leading to loss of Shp1 function both reduces the activity of protein phosphatase 1 (PP1) and suppresses the lethality caused by overexpression of PP1 (ref. 16), indicating that PP1 is activated by Shp1. This effect is probably indirect because Shp1 does not appear to bind PP1 (refs 17, 18). We propose that this effect is mediated through binding of Shp1 to the yeast p97 homologue Cdc48p (ref. 19).

The dependence on p47 of p97 function was assessed using a cell-free

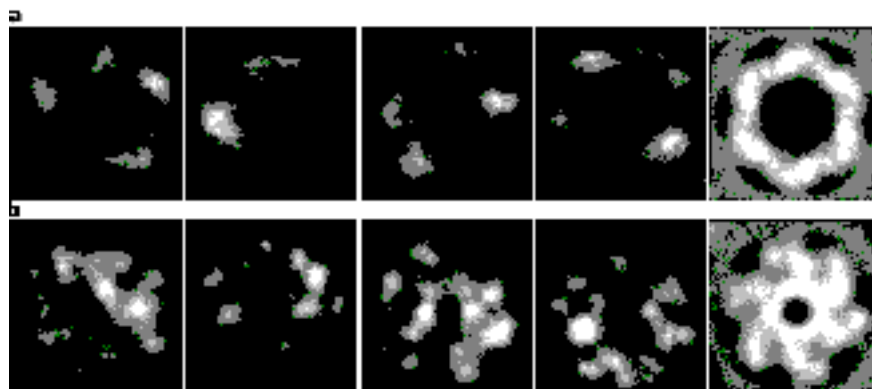


Figure 4 Negative staining of pure p97 and the p97/p47 complex. Negatively stained images of **a**, individual p97 hexamers; and **b**, p97/His₆-p47 complexes. A set of four end-on views are shown on the left, together with a 6-fold 'symmetrized' view on the right. Note the decrease in diameter of the central channel from ~3.5 to 1.2 nm upon incubation with His₆-p47.

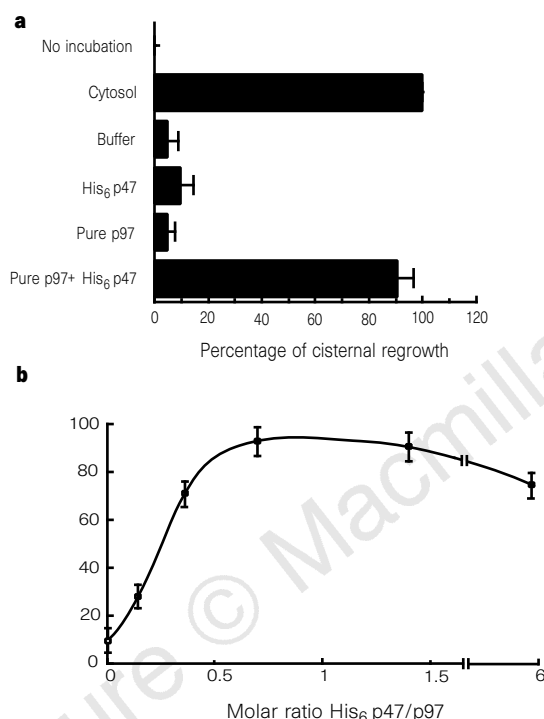


Figure 5 Effect of p47 on p97-mediated cisternal regrowth from mitotic Golgi fragments. NEM-treated mitotic Golgi fragments were incubated with the compounds shown. Saturating levels of His₆-p47 (0.7 to 1.4 mol His₆-p47/mol p97) were used in **a**, and the full range in **b**. Samples were then processed for microscopy and stereology. Results are expressed as a percentage ($\pm$ s.d.) of the regrowth observed using cytosol, where 0% corresponds to $21 \pm 3\%$ of membrane in cisternae and 100% corresponds to $42 \pm 3\%$.

system that mimics the reassembly of the Golgi apparatus at the end of mitosis²⁰. Purified rat liver Golgi stacks were incubated with mitotic cytosol from HeLa cells and the mitotic membrane fragments isolated. After pretreatment with *N*-ethylmaleimide (NEM; to inactivate endogenous p97), these fragments re-form single cisternae when p97 is added⁸. However, our earlier preparations of p97 from rat and *Xenopus* were always contaminated with some p47 (ref. 8). Pure rat p97, lacking p47, was unable to restore growth to NEM-treated fragments (Fig. 5a), as was pure His₆-p47. However, a mixture of pure His₆-p47 and rat p97 restored growth to almost the same level as interphase cytosol (Fig. 5a). Furthermore, varying the ratio of His₆-p47 to p97 showed that maximum regrowth occurred at about 0.7 mol His₆-p47 per 1 mol p97 (Fig. 5b), consistent with the proposed composition of the p97/p47 complex. His₆-p47 had a similar effect on purified *Xenopus* p97

(data not shown). These results indicate that the p97/p47 complex is probably the active component in the fusion reaction that leads to cisternal regrowth. We are now focusing on the role played by p47 in this reaction. □

Methods

Purification of p97 and p47. p97 purified from rat liver cytosol contained variable amounts of p47 (ref. 8). These were separated on a Superose-6 column (Pharmacia FPLC) in TAMD buffer (20 mM Tris-Cl, 1 mM Mg-ATP, 1 mM DTT, pH 7.4) containing 5% glycerol and 1 M KCl.

Relative molecular masses. M_r estimates can be derived from gel filtration if the markers used and the protein being assayed are globular. This is true for p97 (M_r by gel filtration, $570 \pm 10K$ (ref. 13); M_r from sedimentation, $612 \pm 8.5K$ (ref. 12); and M_r from scanning transmission electron microscopy, $590 \pm 5.5K$ (ref. 12)). The axial ratio is also close to unity ($12 \text{ nm diameter}/10.5 \text{ nm length} = 1.14$)¹². M_s s were determined using a Superose-6 column (Pharmacia SMART system) equilibrated in TAMD buffer containing 0.15 M KCl. M_r markers were thyroglobulin (1,340K and 670K), apoferritin (440K), β -amylase (200K) and bovine serum albumin (66K). Fractions were assayed by western blotting using monoclonal anti-p97 and polyclonal anti-p47 (see below). Horseradish peroxidase (HRP)-conjugated second antibodies (Biosource International) were visualized using ECL (Amersham). Blots were scanned and analysed using NIH image software.

Peptide sequence. After SDS-PAGE, p47 was electroblotted onto Immobilon-P PVDF membrane and digested *in situ* with trypsin²¹. Digested peptides were eluted with formic acid:ethanol (1:1 v/v) and purified by reverse-phase HPLC. Eight individual peptides were sequenced using a combination of solid-phase Edman degradation²² and low-energy collision-activated dissociation using a Finnigan MAT TSQ7000 triple-stage quadrupole mass spectrometer²³.

Antibodies. Peptide-5 (LGAPEEESAYVAGER) was conjugated to BSA through an added N-terminal cysteine and used to raise antibodies in rabbits. Monoclonal anti-His antibody was purchased from QIAGEN. Immunoprecipitations were performed using Dynabeads (Dyna) following the manufacturer's protocol.

Molecular cloning. The N- and C-terminal sequences (FIVDAR and MTFPNK, respectively) of the longest peptide (peptide 7, 23 residues) were used to design sense and antisense degenerate primers which were used for PCR amplification of a rat liver cDNA library. The PCR products were cloned into pBluescript II KS+ (Stratagene) and clones of the appropriate size range were sequenced. The cDNA corresponding to peptide 7 was then extended with the Marathon cDNA Amplification Kit (Clontech) using gene specific primers for 5'- and 3'-RACE (5'-RACE: 5'-GTCATAAGGACAAAAGTGGTGGCAG-3'; 3'-RACE: 5'-ATCGTGGATGCACGGCCGGCTATGG-3'). The resulting 5' (1,025 bp) and 3' (407 bp) sequences were used to screen the pUX1 cDNA library from rat liver. Five clones were isolated and sequenced. The missing 5' end was obtained by 5' RACE (gene-specific primer: 5-GCTTGCAAGTG-CAATCTGCAGGTCC-3'; nested primer: 5'-CAGCCAGCTGACTCCAAGAA-GAAAC-3'). Seven clones covered this end and all had identical sequences. Finally, the entire cDNA of 1,488 bp was resequenced; the nucleotide sequence

will appear in the DDBJ, EMBL and GenBank databases under the accession number AB002086.

Northern blotting utilized an MTN blot (Clontech) containing poly(A)⁺ RNA from eight rat tissues (2 µg per lane resolved on a denaturing formaldehyde-agarose gel and transferred to nylon membrane). The full-length cDNA probe was prepared by hexamer-primed labelling (Boehringer Mannheim).

Preparation of His₆-p47. The coding sequence of p47 was amplified without the initiator ATG but with flanking *Bam*HI and *Sall* sites and cloned into pQE-30 (QIAGEN). The expressed protein was purified using Ni-NTA-agarose (Invitrogen) and high-performance Q (Pharmacia).

Light scattering. Measurements were carried out using a miniDAWN machine, following the manufacturer's instructions.

Electron microscopy. Samples were adsorbed on to carbon-coated Formvar grids and negatively stained with 1% uranyl acetate before air-drying. Grids were viewed using a JEOL 1010 electron microscope at 80 kV with a 70 µm objective aperture. Selected images were aligned and filtered using SEMPER software²⁴.

Reassembly assay. NEM-treated mitotic Golgi fragments (10–20 µg)⁸ were incubated with (final concentrations): rat liver cytosol (30% ammonium sulphate cut; 7.9 mg ml⁻¹); pure p97 (20 µg ml⁻¹; ~220 nM); His₆-p47 (15 µg ml⁻¹; ~320 nM); a mixture of His₆-p47 and p97 (p97: 20 µg ml⁻¹; ~220 nM; His₆-p47: 0 to 60 µg ml⁻¹, 0 to ~1.3 µM), preincubated for 30 min on ice before addition to the assay. After incubation for 60 min at 37 °C, samples were processed for electron microscopy and the percentage of membrane in cisternae estimated⁸.

Received 11 March; accepted 30 April 1997.

- Rothman, J. E. & Warren, G. Implications of the SNARE hypothesis for intracellular membrane topology and dynamics. *Curr. Biol.* **4**, 220–233 (1994).
- Rothman, J. E. & Wieland, F. T. Protein sorting by transport vesicles. *Science* **272**, 227–234 (1996).
- Pryer, N. K., Wuestehube, L. J. & Schekman, R. Vesicle-mediated protein sorting. *Annu. Rev. Biochem.* **61**, 471–516 (1992).
- Ferro-Novick, S. & Jahn, R. Vesicle fusion from yeast to man. *Nature* **370**, 191–193 (1994).
- Nuoffer, C. & Balch, W. E. GTPases—multifunctional molecular switches regulating vesicular traffic. *Annu. Rev. Biochem.* **63**, 949–990 (1994).
- Schiavo, G., Rossetto, O. & Montecucco, C. Clostridial neurotoxins as tools to investigate the molecular events of neurotransmitter release. *Sem. Cell Biol.* **5**, 221–229 (1994).
- Pfeffer, S. R. Transport vesicle docking: SNAREs and associates. *Annu. Rev. Cell Dev. Biol.* **12**, 441–461 (1996).
- Rabouille, C., Levine, T. P., Peters, J. M. & Warren, G. An NSF-like ATPase, p97, and NSF mediate cisternal regrowth from mitotic Golgi fragment. *Cell* **82**, 905–914 (1995).
- Acharya, U. *et al.* The formation of Golgi stacks from vesiculated Golgi membranes requires two distinct fusion events. *Cell* **82**, 895–904 (1995).
- Latterich, M., Frohlich, K. U. & Schekman, R. Membrane fusion and the cell cycle: Cdc48p participates in the fusion of ER membranes. *Cell* **82**, 885–893 (1995).
- Warren, G. Membrane partitioning during cell division. *Annu. Rev. Biochem.* **62**, 323–348 (1993).
- Peters, J. M. *et al.* Ubiquitous soluble Mg²⁺-ATPase complex: A structural study. *J. Mol. Biol.* **223**, 557–571 (1992).
- Peters, J. M. & Walsh, M. J. & Franke, W. W. An abundant and ubiquitous homo-oligomeric ring-shaped ATPase particle related to the putative vesicle fusion proteins, Sec18p and NSF. *EMBO J.* **9**, 1757–1767 (1990).
- Confalonieri, F., Marsault, J. & Duguet, M. SAV, an archaeobacterial gene with extensive homology to a family of highly conserved eukaryotic ATPases. *J. Mol. Biol.* **235**, 396–401 (1994).
- Secheren, B., el Bakkoury, M., Vierendeels, F., Dubois, E. & Messenguy, F. Sequencing and functional analysis of a 32,560 bp segment on the left arm of yeast chromosome II. Identification of 26 open reading frames, including the KIP1 and SEC17 genes. *Yeast* **9**, 1355–1371 (1993).
- Zhang, S., Guha, S. & Volkert, F. C. The *Saccharomyces SHP1* gene, which encodes a regulator of phosphoprotein phosphatase 1 with differential effects on glycogen metabolism, meiotic differentiation, and mitotic cell cycle progression. *Mol. Cell. Biol.* **15**, 2037–2050 (1995).
- Tu, J. L., Song, W. J. & Carlson, M. Protein phosphatase type I interacts with proteins required for meiosis and other cellular processes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **16**, 4199–4206 (1996).
- Sutton, A., Lin, F. & Arndt, K. T. The SIT4 protein phosphatase is required in late G1 for progression into S phase. *Cold Spring Harb. Symp. Quant. Biol.* **56**, 75–81 (1991).
- Frohlich, K.-U. *et al.* Yeast cell cycle protein CDC48p shows full-length homology to the mammalian protein VCP and is a member of a protein family involved in secretion, peroxisome formation, and gene expression. *J. Cell Biol.* **114**, 443–453 (1991).
- Rabouille, C., Misteli, T., Watson, R. & Warren, G. Reassembly of Golgi stacks from mitotic Golgi fragments in a cell-free system. *J. Cell Biol.* **129**, 605–618 (1995).
- Sutton, C. W. *et al.* Identification of myocardial proteins from two-dimensional gels by peptide mass fingerprinting. *Electrophoresis* **16**, 308–316 (1995).
- Coull, J. M., Pappin, D. J., Mark, J., Aebersold, R. & Koster, H. Functionalized membrane supports for covalent protein microsequence analysis. *Analyt. Biochem.* **194**, 110–120 (1991).
- Hunt, D. F., Yates, J. R., Shabanowitz, J., Winston, S. & Hauer, C. R. Protein sequencing by tandem mass spectrometry. *Proc. Natl Acad. Sci. USA* **83**, 6233–6237 (1986).
- Saxton, W. O. Sempers: distortion compensation, selective averaging, 3-D reconstruction, and transfer function correction in a highly programmable system. *J. Struct. Biol.* **116**, 230–236 (1996).

Acknowledgements. We thank T. Chappell, F. Barr, B. Svejstrup, B. Sönnichsen, M. Lowe, C. Fernandez, D. Shima, J. Shorter, N. Hui and A. Coffer for helpful comments and for reagents; G. Banting for the rat liver cDNA library; J.-M. Peters for monoclonal anti-p97 antibodies; G. Clark and A. Davies for DNA sequencing; and D. Rahman for protein sequencing. Special thanks go to N. Nakamura for advice.

Correspondence and requests for materials should be addressed to G.W. (e-mail: g.warren@europa.lif.icnet.uk).

Cofilin promotes rapid actin filament turnover *in vivo*

Pekka Lappalainen & David G. Drubin

Department of Molecular and Cell Biology, 401 Barker Hall, University of California, Berkeley, California 94720, USA

The ability of actin filaments to function in cell morphogenesis and motility is coupled to their capacity for rapid assembly and disassembly. Because disassembly *in vitro* is much slower than *in vivo*, cellular factors that stimulate disassembly have long been assumed to exist. Although numerous proteins can affect actin dynamics *in vitro*, demonstration of *in vivo* relevance of these effects has not been achieved. We have used genetics and an actin-inhibitor in yeast to demonstrate that rapid cycles of actin assembly and disassembly depend on the small actin-binding protein cofilin, and that cofilin stimulates filament disassembly. These results may explain why cofilin is ubiquitous in eukaryotes and is essential for viability in every organism in which its function has been tested genetically. Magnitudes of disassembly defects in cofilin mutants *in vivo* were found to be correlated closely with the magnitudes of disassembly defects observed *in vitro*, supporting our conclusions. Furthermore, these cofilin

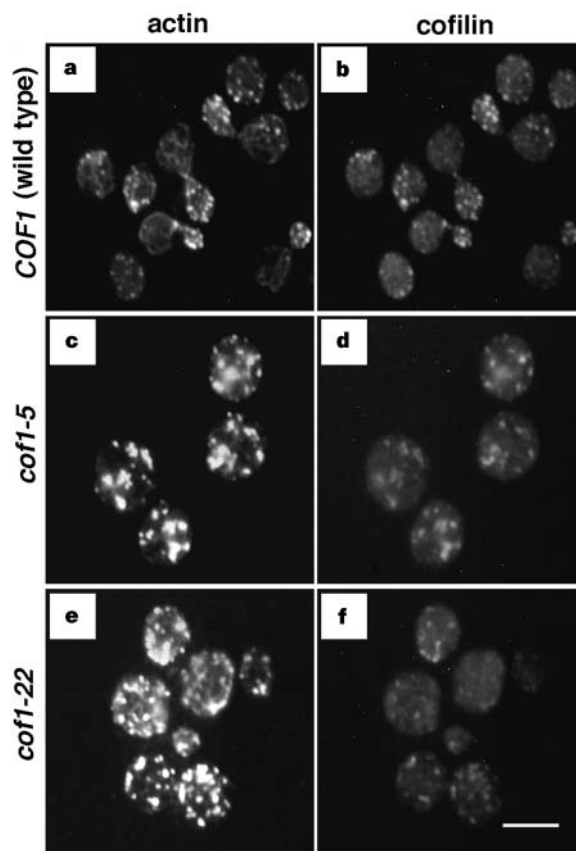


Figure 1 Cells were grown at 20 °C to $A_{600} = 0.3$, shifted to 34 °C for 3 h and fixed with 5% formaldehyde. The actin and cofilin in wild-type (a, b), *cof1-5* (c, d) and *cof1-22* (e, f) were visualized by immunofluorescence. In the wild-type cells, the actin patches are well polarized and actin cables are visible in the cells that are displaying polarized growth. Cofilin localizes to cortical actin patches. In *cof1-5* and *cof1-22* cells shifted to 34 °C the actin patches are very large, irregularly shaped and almost completely depolarized. In both mutants, cofilin localizes mostly to these actin structures. Scale bar, 5 µm.

mutants provided an opportunity to distinguish in living cells those actin functions that depend specifically on filament turnover (endocytosis) from those that do not (cortical actin patch motility).

Cofilin is a member of the cofilin/ADF family of small (M_r 15K–20K) actin-binding proteins¹. These proteins bind to actin monomers (G-actin) and actin filaments (F-actin) *in vitro*, and stimulate depolymerization of actin filaments in a pH-dependent manner^{2–5}. Despite a wealth of biochemical data, little is known about the function of cofilin and related proteins *in vivo*. Every eukaryotic cell type examined contains at least one member of this family, suggesting that cofilin performs a universally important function¹. This conclusion is further supported by the observations that in *Saccharomyces cerevisiae*, *Drosophila melanogaster* and *Caenorhabditis elegans*, null mutations in the cofilin/ADF genes are lethal^{6–8}. However, the lack of partial loss-of-function mutations in cofilin had previously precluded tests of the roles of this protein in living cells.

We have studied the contribution of cofilin to actin cytoskeleton dynamics in yeast by using two temperature-sensitive cofilin alleles that were isolated in a systematic mutational analysis of the *S. cerevisiae* *COF1* gene (P.L. *et al.*, unpublished results). In both alleles, a cluster of charged residues (Asp 10 and Glu 11 in *cof1-5*, and Glu 134, Arg 135 and Arg 138 in *cof1-22*) are replaced by alanines. The mutant alleles were integrated into the yeast genome and their phenotypes examined in haploid segregants. At temperatures suitable for robust growth (20–25°C), *cof1-5* and *cof1-22* strains show similar growth rates and morphology to wild-type cells. The actin 'patches', which are F-actin structures associated with the cell cortex⁹, are slightly brighter in *cof1-5* and *cof1-22* cells than in wild-type cells, but overall organization of the cytoskeleton seems to be normal, and actin levels in these mutants are unchanged within the sensitivity of quantitative western blots ($\pm 20\%$). At a temperature that was non-permissive for growth

(34°C), *cof1-5* and *cof1-22* cells became enlarged and their cortical actin patches lost their characteristic asymmetric localization (Fig. 1). After incubation at 34°C for 3 h, 15–20% of the mutant cells have multiple nuclei. The actin patches in both *cof1-5* and *cof1-22* cells are unusually large (Fig. 1), consistent with the possibility that these cells have severe defects in actin filament turnover.

To examine the effects of these mutations on F-actin turnover rates *in vivo*, we used latrunculin-A (Lat-A), an actin monomer-sequestering drug. Lat-A binds close to the nucleotide binding pocket of actin monomers, and prevents actin filament assembly while allowing disassembly^{10,11}. In living yeast cells Lat-A causes a rapid and specific disruption of the actin cytoskeleton¹¹. Because Lat-A functions by sequestering actin monomers, the rate of disappearance of actin structures is expected to reflect the rate of subunit dissociation from actin filaments *in vivo*¹¹. The Lat-A actin-filament disassembly assay was performed at both 25°C and 34°C. The rates of disassembly in wild-type and cofilin mutant strains seem to be independent of temperature. The data from the depolymerization assay at 25°C are presented in Figs 2 and 3a. In wild-type cells, most of the actin patches disappear within 2 min of addition of a high concentration (400 μ M) of Lat-A. After 5 min only ~30% of the wild-type cells have any visible actin-filament structures when visualized by rhodamine-phalloidin staining (Figs 2 and 3a). The proportion of cells with faint actin-filament staining remains unchanged at 10 min. The rate of disappearance of actin patches in *cof1-5* cells is slightly reduced compared with wild-type cells, whereas the patches in *cof1-22* cells disappear much more slowly (Figs 2 and 3a). Most of the *cof1-22* cells still had many bright actin patches 10 min after the addition of Lat-A, indicating that this cofilin mutation causes a dramatic decrease in the F-actin depolymerization rate *in vivo* (the intensity and number of actin patches in Fig. 2i (*cof1-22*, 10 min) is greater than in Fig. 2b (*COF1*, 2 min)). The total disappearance of actin filament structures in *cof1-22* cells takes approximately 20 min (data not shown).

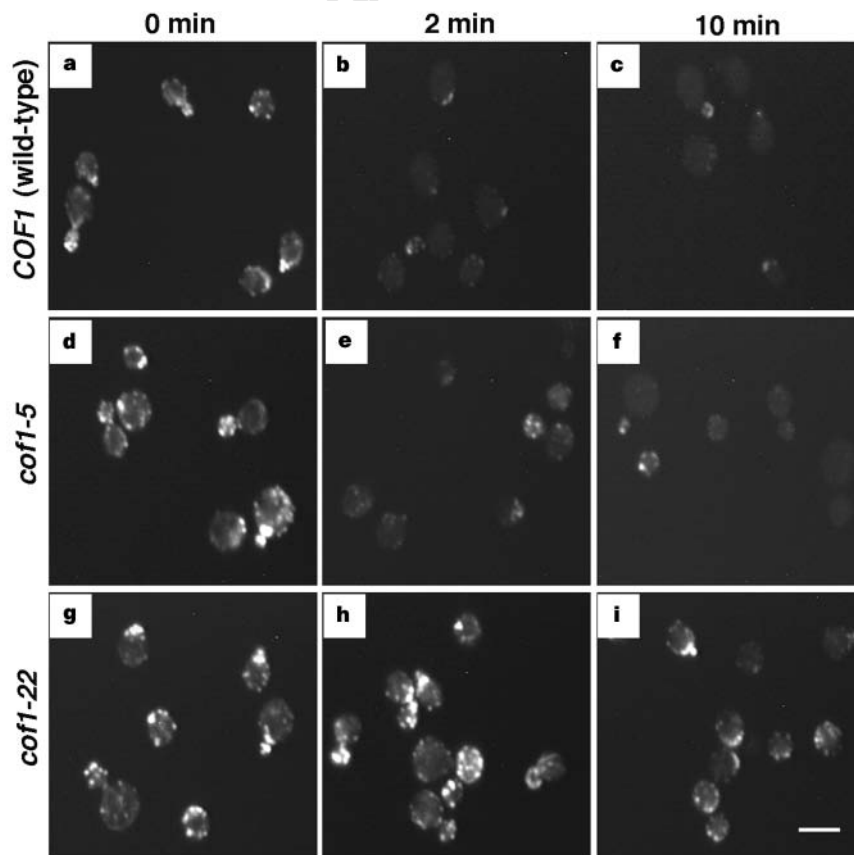


Figure 2 Filamentous actin was visualized by rhodamine-phalloidin staining in wild-type (a–c), *cof1-5* (d–f) and *cof1-22* (g–i) strains at 0, 2 and 10 min after the addition of 400 μ M Lat-A to the cultures at 25°C. In wild-type and *cof1-5* cells, the actin filaments disappear rapidly. After 2 min cells have either very faint actin patches or none at all. In contrast, actin patches in *cof1-22* cells disappear >5-fold more slowly than in the wild type. Even after 10 min, most of the *cof1-22* cells have brightly stained actin patches. All photographs were taken using identical microscope and imaging-system settings. Scale bar, 5 μ m.

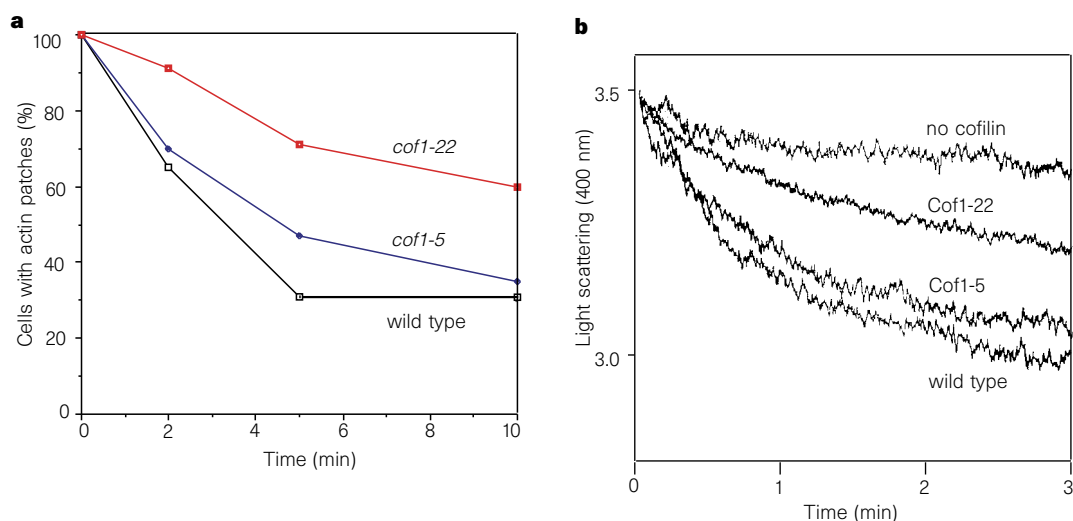


Figure 3 The effects of coflin mutations on actin-filament depolymerization *in vivo* and *in vitro*. **a**, The percentage of cells with visible actin-filament structures in wild-type, *cof1-5* and *cof1-22* cells at different times after the addition of Lat-A at 25°C was quantified by rhodamine-phalloidin staining and fluorescent microscopy. For each strain and time point, at least 200 cells from two independent experiments were scored for the presence of actin patches. **b**, The F-actin depolymerization stimulated by recombinant wild-type, *Cof1-5* and *Cof1-22*

coflin was followed by the decrease in light-scattering at 400 nm. Note that dilution of 5 μ M yeast actin filaments to 0.5 μ M in the absence of coflin results in slow actin-filament depolymerization. Wild-type coflin and *Cof1-5* cause large increases in the rates of depolymerization, whereas recombinant *Cof1-22* causes only a small increase in actin-filament depolymerization rates. The coflin and yeast actin samples used in these experiments were more than 99% and 95% pure, respectively, based on Coomassie-stained SDS gels (data not shown).

The relative rates of F-actin depolymerization measured *in vivo* for the coflin mutants correlate well with the rates by which they promote F-actin depolymerization *in vitro*. Recombinant *Cof1-5* and *Cof1-22* mutants were expressed and purified. In fluorescence-monitored urea denaturation assays, both mutants show simple two-state unfolding transitions, indicating that the preparations consist of homogeneous populations of protein (P.L. *et al.*, unpublished results). To assay depolymerization rates *in vitro*, actin filaments were first assembled from purified yeast actin. They were then diluted 10-fold with or without recombinant wild-type, *Cof1-5* or *Cof1-22* yeast coflin. The depolymerization of actin filaments was followed by a decrease in light scattering at 400 nm. In this *in vitro* assay, performed at 25°C and 37°C, *Cof1-5* showed a small defect in actin-filament depolymerization, and *Cof1-22* showed a pronounced defect in actin-filament depolymerization relative to wild-type coflin (Fig. 3b). The magnitudes of these defects were not affected by temperature. Because *Cof1-5* showed only a very small depolymerization defect in our *in vivo* and *in vitro* assays, it is possible that the temperature sensitivity of this mutant results from some currently unidentified activity of coflin, or that the small decrease in actin dynamics precludes cell growth at elevated temperatures. When the *in vitro* depolymerization assay was performed by adding 100 μ M Lat-A and 1 μ M coflin to 5 μ M actin filaments, results similar to those in Fig. 3b were obtained, indicating that Lat-A does not affect the rate of coflin-stimulated actin depolymerization (data not shown).

To test the specificity of the effects for coflin, we applied the Lat-A assay to yeast strains carrying deletions in genes that encode other proteins important for actin cytoskeleton organization (*SLA2*, *SAC6* and *SRV2*)¹²⁻¹⁴. In these strains, disappearance of actin patches follows kinetics similar to those observed for their wild-type parent strains (data not shown). As *sac6* and *sla2* mutants are defective in endocytosis, these results demonstrate that endocytosis defects do not impair the entry of Lat-A into yeast cells. Further, *cof1-5* and *cof1-22* strains showed fivefold and fourfold greater sensitivity to Lat-A, respectively, than the wild type, further suggesting that these strains do not have defects in Lat-A uptake (data not shown).

The availability of coflin mutants that cause a dramatic decrease

in actin-filament turnover provided an opportunity to determine the role of actin-filament turnover in living cells. Many components in the yeast cortical actin cytoskeleton have been shown to be important for endocytosis^{13,15,16}. To test specifically whether the F-actin turnover induced by coflin is necessary for endocytosis, we followed the uptake of a fluid-phase marker, lucifer yellow, in *cof1-5* and *cof1-22* strains. At temperatures permissive for robust cell growth (20 and 25°C), *cof1-5* and *cof1-22* cells show severe defects in the uptake of lucifer yellow (Fig. 4a), indicating that rapid turnover of cortical actin structures is essential for endocytosis. Consistent with this possibility, the coflin mutant with the most dramatic effects on actin turnover, *cof1-22*, had the most dramatic effects on endocytosis. It is possible that the temperature-sensitive phenotype of *cof1-5* and *cof1-22* mutants results from the defect in endocytosis, as an apparent dependence on endocytosis for growth at high temperatures has been reported¹⁷. Alternatively, cells may have a higher requirement for F-actin depolymerization at higher temperatures, or there could be another coflin-dependent reaction that is sensitive to temperature.

Yeast cortical actin patches have recently been shown to translocate within the plane of the cell cortex with a velocity up to 1.5 μ m s⁻¹ (refs 18, 19). As these studies failed to demonstrate a role for any known myosins in this motility¹⁸, it has been suggested that the actin-patch movement might be driven by a filament turnover mechanism similar to that which drives intracellular movement of *Listeria monocytogenes*²⁰. To test whether actin-filament turnover is required for this motility, we transformed wild-type and *cof1-22* mutant strains with a plasmid expressing a green fluorescent protein (GFP)-Abp1 fusion protein that localizes specifically to cortical actin structures¹⁹. At permissive temperature (20–25°C) the cortical actin patches give strong fluorescence and show rapid movements (Fig. 4b). The quantified data from approximately 700 actin-patch movements in wild-type and *cof1-22* mutant cells is shown in Fig. 4c. There are no significant differences either in the speed distribution or in the observed maximum velocity of actin patches between wild type and the *cof1-22* mutant strain. Because the *cof1-22* mutant shows a large decrease in actin-filament turnover *in vivo*, these results suggest that

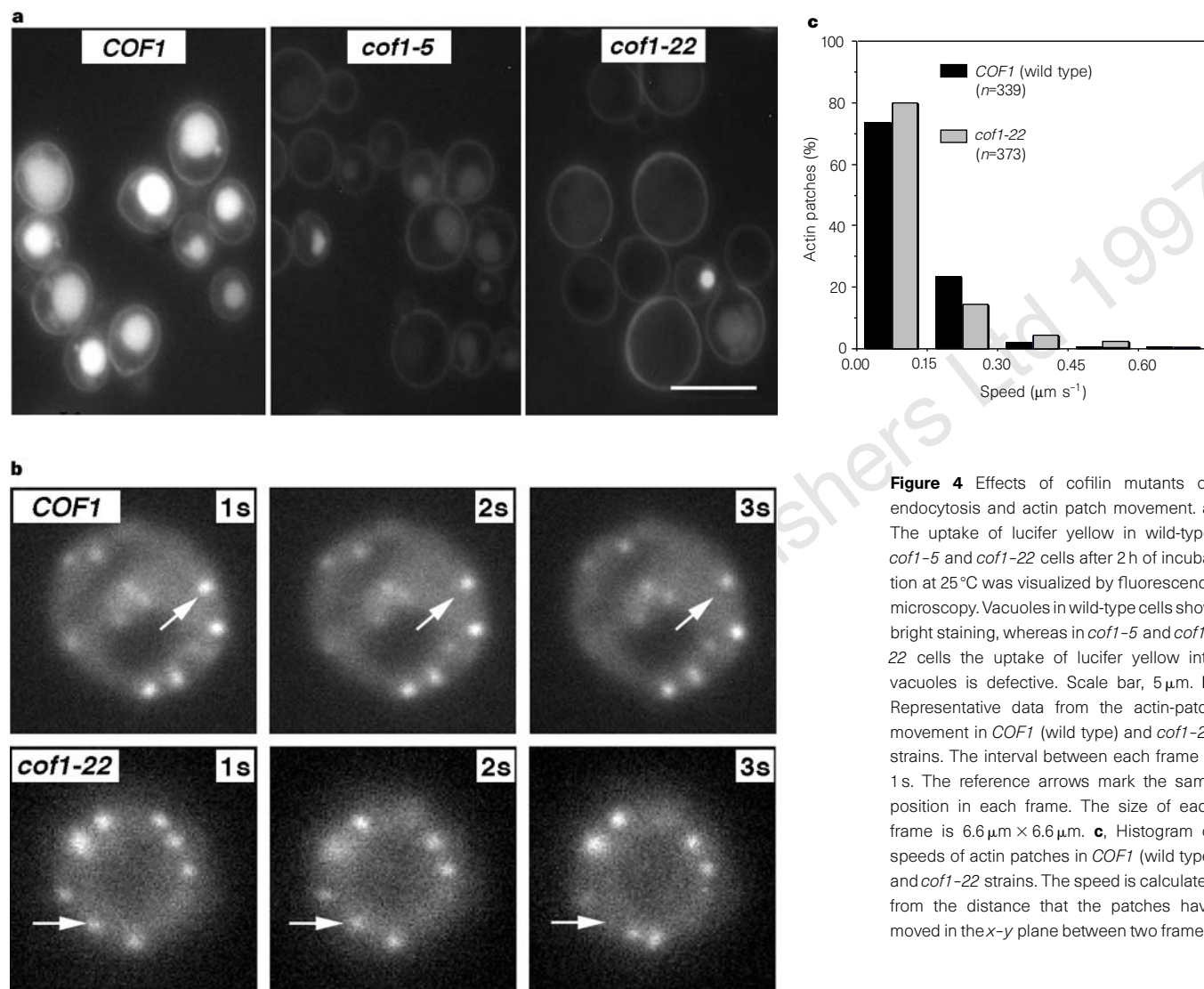


Figure 4 Effects of cofilin mutants on endocytosis and actin patch movement. **a**, The uptake of lucifer yellow in wild-type, *cof1-5* and *cof1-22* cells after 2 h of incubation at 25 °C was visualized by fluorescence microscopy. Vacuoles in wild-type cells show bright staining, whereas in *cof1-5* and *cof1-22* cells the uptake of lucifer yellow into vacuoles is defective. Scale bar, 5 μm . **b**, Representative data from the actin-patch movement in *COF1* (wild type) and *cof1-22* strains. The interval between each frame is 1 s. The reference arrows mark the same position in each frame. The size of each frame is 6.6 $\mu\text{m} \times 6.6 \mu\text{m}$. **c**, Histogram of speeds of actin patches in *COF1* (wild type) and *cof1-22* strains. The speed is calculated from the distance that the patches have moved in the x-y plane between two frames.

actin-patch motility is not driven by filament turnover. In support of this conclusion, we find that patch motility is unaffected by sequestration of free actin monomers through the addition of high concentrations of Lat-A to cells (P.L. and L. Belmont, unpublished data). Patches continued to move even as they were disassembling.

In conclusion, we have provided evidence that members of the cofilin/ADF family promote actin-filament depolymerization *in vivo* and are required for rapid cycles of actin assembly and disassembly. It has been shown that cofilin/ADF promotes actin-filament depolymerization in cell extracts^{21,22}. Our data extend this observation to living cells, and identify a process in which rapid turnover of actin filaments is essential. These findings, considered with the intimate coupling of actin dynamics with actin function, may explain why cofilin is expressed ubiquitously and is essential for viability in every cell type in which its gene has been mutated. The actin-associated activities of cofilin have been shown to be regulated by pH, phosphorylation and phosphoinositides^{3,23-25}. However, as we were unable to detect any differences in Lat-A-induced disassembly rates among cells at different stages of the cell cycle (data not shown), our data indicate that in yeast, cofilin is active throughout the cell cycle. Despite speculation that cortical actin-patch motility might be driven by an assembly coupled mechanism akin to that in *Listeria* bacteria²⁰, loss of cofilin function did not have consequences for cortical actin-patch motility. Considered with a previous study that failed to find an effect of mutation of any of the

yeast myosin genes on patch motility¹⁸, our findings suggest that patch motility might be driven by a different mechanism. □

Methods

Yeast strains. The construction of *cof1* mutants will be described in more detail elsewhere (P.L. *et al.*, unpublished data). Briefly, the *LEU2* marker was inserted 91 base pairs downstream from the end of the cofilin open reading frame (ORF) in pBluescript SK plasmid (Stratagene) carrying *COF1*. The site-directed mutations were created by oligonucleotide-based mutagenesis (Transformer, Clontech). The plasmids were linearized and *COF1*, *cof1-5* and *cof1-22* were integrated into a *COF1/cof1-Δ1::HIS3* hemizygote strain (DDY427). Integrants were identified by screening for loss of the *HIS3* marker and gain of the *LEU2* marker. These cells were sporulated and the phenotypes were scored in haploid segregants. Each mutation was confirmed by PCR amplification and restriction endonuclease digestions. The haploid yeast strains carrying deletions of *SLA2*, *SAC6* and *SRV2* and their parent strains have been described^{12,26,27}.

Cell biological methods. Yeast cultures were grown in YPD medium at 20 °C, unless stated otherwise. Immunofluorescence was performed as described⁶. The fluid-phase endocytosis assay was performed at 25 °C as described²⁸. For the *in vivo* Lat-A depolymerization assay, cells were grown at 20 °C at $A_{600} = 0.2$, and Lat-A was added to cultures to a final concentration of 400 μM . The cells were incubated at 25 or 34 °C and 250- μl samples of cells were fixed at 0, 2 and 10 min with 5% formaldehyde, and actin-filament structures were visualized by rhodamine-phalloidin staining. After staining for

30 min, the cells were washed 5 times with 200 μ l PBS + 0.1% TritonX-100, resuspended in 6 μ l mounting media and placed on a slide. The sensitivities of wild-type, *cof1-5* and *cof1-22* cells to lat-A were examined in a halo-assay¹¹ in which filters spotted with Lat-A were deposited on lawns of yeast and the diameter of zones of growth inhibition measured. For actin-patch movement studies, the *COF1* and *cof1-22* strains were transformed with a plasmid carrying GFP linked to Abp1 (ref. 19). Cells from such transformants were grown at 20 °C to an A_{600} of ~0.4, and movement of actin patches was followed with Zeiss Axiovert 100TV microscope. Images from about 400 patches in 5 unbudded cells from each strain were collected using a Princeton Instruments ST-138 camera/controller with a Kodak KAF1400 chip (each pixel is 0.069 μ m $\times$ 0.069 μ m). Frames from the cells were recorded at intervals of 0.95 to 1.25 s and the data analysed using the Animate program (written by A. Mallavarapu). The maximum speeds observed in our study (0.75 μ m s⁻¹) are approximately twofold slower than the maximum speeds previously reported¹⁸. This might result from differences in yeast strain backgrounds or the differences in the time resolution of image acquisition (the camera system in ref. 18 recorded frames with 0.2-s intervals compared to ~1-s intervals in this study).

In vitro depolymerization assay. The *COF1* yeast cofilin gene was amplified by PCR, subcloned into pGEX2T vector (Pharmacia), and expressed in *Escherichia coli* JM109 cells as a glutathione S-transferase (GST) fusion protein. Site-directed mutations were introduced to pGEX2T-*COF1* plasmid by oligonucleotide-based mutagenesis (Transformer, Clontech). Fusion proteins were purified by glutathione-agarose affinity chromatography²⁹. All constructs made by PCR were sequenced to exclude the possibility of undesired mutations. Cofilin was cleaved from GST by thrombin and purified by gel-filtration (Superdex-75, Pharmacia). Yeast actin was purified as described³⁰ and polymerized in F-buffer comprising (in mM) 20 Tris, pH 7.5, 100 KCl, 2 MgCl₂, 0.7 ATP, 0.2 DTT, 0.2 CaCl₂, at a final concentration of 5 μ M. For the depolymerization assay, F-actin and cofilin were mixed in a cuvette to final concentrations of 0.5 μ M, and the depolymerization of F-actin was followed by decrease in light scattering at 400 nm.

Received 13 March; accepted 21 April 1997.

1. Moon, A. & Drubin, D. G. The ADF-cofilin proteins: Stimulus responsive modulators of actin dynamics. *Mol. Biol. Cell* **6**, 1423–1431 (1995).
2. Bamburg, J. R., Harris, H. E. & Weeds, A. G. Partial purification and characterization of an actin depolymerizing factor from brain. *FEBS Lett.* **121**, 178–182 (1990).
3. Yonezawa, N., Nishida, E. & Sakai, H. pH control of actin polymerization by cofilin. *J. Biol. Chem.* **260**, 14410–14412 (1985).
4. Hawkins, M., Pope, B., Maciver, S. K. & Weeds, A. G. Human actin depolymerizing factor mediates a pH-sensitive destruction of actin filaments. *Biochemistry* **32**, 9985–9993 (1993).
5. Hayden, S. M., Miller, P. S., Brauweiler, A. & Bamburg, J. R. Analysis of the interactions of actin depolymerizing factor with G- and F-actin. *Biochemistry* **32**, 9994–10004 (1993).
6. Moon, A. L., Janmey, P. A., Louie, K. A. & Drubin, D. G. Cofilin is an essential component of the yeast cortical cytoskeleton. *J. Cell Biol.* **120**, 421–435 (1993).
7. Gunsalus, K. C. *et al.* Mutations in *twinstar*, a *Drosophila* gene encoding a cofilin-ADF homologue, result in defects in centrosome migration and cytokinesis. *J. Cell Biol.* **131**, 1243–1259 (1995).
8. McKim, K. S., Matheson, C., Marra, M. A., Wakarchuk, M. F. & Baillie, D. L. The *Caenorhabditis elegans unc-60* gene encodes protein homologous to a family of actin binding proteins. *Mol. Gen. Genet.* **242**, 346–357 (1994).
9. Mulholland, J. *et al.* Ultrastructure of the yeast actin cytoskeleton and its association with the plasma membrane. *J. Cell Biol.* **125**, 381–391 (1994).
10. Coue, M., Brenner, S. L., Spector, I. & Korn, E. D. Inhibition of actin polymerization by Latrunculin-A. *FEBS Lett.* **213**, 316–318 (1987).
11. Ayscough, K. R. *et al.* High rates of actin filament turnover in budding yeast and roles for actin in establishment and maintenance of cell polarity revealed using the actin inhibitor Latrunculin-A. *J. Cell Biol.* **137**, 399–416 (1997).
12. Holtzman, D. A., Yang, S. & Drubin, D. G. Synthetic-lethal interactions identify two novel genes, *SLA1* and *SLA2*, that control membrane cytoskeleton assembly in *Saccharomyces cerevisiae*. *J. Cell Biol.* **122**, 635–644 (1993).
13. Kubler, E. & Riezman, H. Actin and fimbrin are required for the internalization step of endocytosis in yeast. *EMBO J.* **12**, 2855–2862 (1993).
14. Freeman, N. L., Chen, Z., Horenstein, J., Weber, A. & Field, J. An actin monomer binding activity localizes to the carboxyl half of the *Saccharomyces cerevisiae* cyclase associated protein. *J. Biol. Chem.* **270**, 5680–5685 (1995).
15. Geli, M. I. & Riezman, H. Role of type I myosins in receptor-mediated endocytosis in yeast. *Science* **272**, 533–535 (1996).
16. Goodson, H. V. *et al.* Synthetic lethality screen identifies a novel yeast myosin I gene (*MYO5*)—Myosin I proteins are required for polarization of the actin cytoskeleton. *J. Cell Biol.* **133**, 1277–1291 (1996).
17. Benedetti, H., Rath, S., Crauzas, F. & Riezman, H. The *END3* gene encodes a protein that is required for the internalization step of endocytosis and for actin cytoskeleton organization in yeast. *Mol. Biol. Cell* **5**, 1023–1037 (1994).
18. Waddle, J. A., Karpova, T. S., Waterston, R. H. & Cooper, J. A. Movement of cortical actin patches in yeast. *J. Cell Biol.* **132**, 861–870 (1996).
19. Doyle, T. & Botstein, D. Movement of yeast cortical actin cytoskeleton visualized *in vivo*. *Proc. Natl Acad. Sci. USA* **93**, 3886–3891 (1996).
20. Welch, M. D., Mallavarapu, A., Rosenblatt, J. & Mitchison, T. J. Actin dynamics *in vivo*. *Opin. Cell Biol.* **9**, 54–61 (1997).

21. Rosenblatt, J., Agnew, B. J., Abe, H., Bamburg, J. R. & Mitchison, T. J. *Xenopus* actin depolymerizing factor/cofilin XAC is responsible for the turnover of actin filaments in *Listeria monocytogenes* tails. *J. Cell Biol.* **136**, 1323–1332 (1997).
22. Carlier, M. *et al.* Actin depolymerizing factor (ADF/cofilin) uses ATP hydrolysis to enhance actin dynamics. *J. Cell Biol.* **136**, 1307–1322 (1997).
23. Yonezawa, N., Nishida, E., Iida, K., Yahara, I. & Sakai, H. Inhibition of the interactions of cofilin, destrin and deoxyribonuclease-I with actin by phosphoinositides. *J. Biol. Chem.* **265**, 8382–8386 (1990).
24. Agnew, B. J., Minamide, L. S. & Bamburg, J. R. Reactivation of phosphorylated actin depolymerizing factor and identification of the regulatory site. *J. Biol. Chem.* **270**, 17582–17585 (1995).
25. Abe, H., Obinata, T., Minamide, L. S. & Bamburg, J. R. *Xenopus laevis* actin-depolymerizing factor/cofilin: A phosphorylation-regulated protein essential for development. *J. Cell Biol.* **132**, 871–885 (1996).
26. Adams, A. E. M., Botstein, D. & Drubin, D. G. Requirement of yeast fimbrin for actin organization and morphogenesis *in vivo*. *Nature* **354**, 404–408 (1991).
27. Lila, T. & Drubin, D. G. Evidence for physical and functional interactions among two *Saccharomyces cerevisiae* SH3 domain proteins, an adenyl cyclase-associated protein and the actin cytoskeleton. *Mol. Biol. Cell* **8**, 367–385 (1997).
28. Dulic, V. *et al.* Yeast endocytosis assays. *Methods Enzymol.* **194**, 697–710 (1991).
29. Ausubel, F. M. *et al.* *Current Protocols in Molecular Biology* (John Wiley, New York, 1990).
30. Buzan, J. M. & Frieden, C. Yeast actin: Polymerization kinetic studies of wild-type and a poorly polymerizing mutant. *Proc. Natl Acad. Sci. USA* **93**, 91–95 (1996).

Acknowledgements. We thank K. Ayscough for the Lat-A disassembly procedure; L. Belmont, B. Goode and K. Kozminski for comments on the manuscript; and A. Mallavarapu and T. Mitchison for help in recording patch movements. This work was supported by long-term fellowships from the European Molecular Biology Organization and Human Frontier Science Program (to P.L.) and by grants from the NIH and American Cancer Society (to D.G.D.).

Correspondence and requests for materials should be addressed to D.G.D. (e-mail: drubin@mendel.berkeley.edu).

Mutations increasing autoinhibition inactivate tumour suppressors Smad2 and Smad4

Akiko Hata*, Roger S. Lo*, David Wotton*, Giorgio Lagna† & Joan Massagué*

* Cell Biology Program and Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, † Laboratory of Molecular Embryology, The Rockefeller University, New York, New York 10021, USA

Smad2 and Smad4 are related tumour-suppressor proteins^{1,2}, which, when stimulated by the growth factor TGF- β , form a complex to inhibit growth³. The effector function of Smad2 and Smad4 is located in the conserved carboxy-terminal domain (C domain) of these proteins and is inhibited by the presence of their amino-terminal domains (N domain)^{4,5}. This inhibitory function of the N domain is shown here to involve an interaction with the C domain that prevents the association of Smad2 with Smad4. This inhibitory function is increased in tumour-derived forms of Smad2 and 4 that carry a missense mutation in a conserved N domain arginine residue. The mutant N domains have an increased affinity for their respective C domains, inhibit the Smad2–Smad4 interaction, and prevent TGF β -induced Smad2–Smad4 association and signalling. Whereas mutations in the C domain disrupt the effector function of the Smad proteins, N-domain arginine mutations inhibit SMAD signalling through a gain of autoinhibitory function. Gain of autoinhibitory function is a new mechanism for inactivating tumour suppressors.

SMAD proteins are central to signalling by the TGF- β family of growth factors⁶. An effector function of the C domain and an inhibitory function of the N domain of Smad proteins were discovered during studies on the mesoderm-inducing activity of Smad2 in *Xenopus* embryo explants⁵ and on the transcriptional activity of various Smad proteins in mammalian cells⁴. SMAD proteins exist as homo-oligomers that form functional hetero-oligomeric complexes in response to TGF- β and related factors³. These complexes are formed by association of a pathway-restricted SMAD with the shared SMAD, Smad4 (also known as DPC4 or deleted in pancreatic carcinoma 4). Thus Smad1 associates with Smad4 in response to stimulation by bone morphogenic protein

(BMP)³, whereas Smad2, or its close isoform Smad3, associates with Smad4 in response to stimulation by TGF- β or activin^{3,7}. These associations occur on receptor-mediated phosphorylation of the pathway-restricted SMADs^{3,7-10}.

To investigate the domains involved in these interactions, we tested various Smad4 fragments as baits either against Smad4, to detect homo-oligomeric interaction, or against Smad2, to detect hetero-oligomeric interaction, in a yeast two-hybrid system. Full-length Smad4 interacted with the N domain/linker region as a whole but not with those two regions when they were separately expressed (Fig. 1a). Full-length Smad4 interacted with its isolated C domain (Fig. 1a). These results indicate that both the C domain and the N domain/linker region can contribute to the homo-oligomeric interaction in Smad4 (Fig. 1a). Furthermore, the isolated C domain of Smad4 interacted with itself (data not shown), which was unexpected¹¹. The homo-oligomeric interaction pattern of

Smad4 in yeast is consistent with a contribution of all three regions to the homo-oligomeric interaction, with the C domain providing the strongest. The crystal structure of the Smad4 C domain has revealed that this domain forms a homotrimer whose interfaces are the targets of cancer mutations¹². Smad2 gave a similar, but not identical, pattern of homo-oligomeric interactions in yeast (Fig. 1a).

The Smad2-Smad4 complex was detectable in yeast and was particularly sensitive to deletions in the C domain (Fig. 1a). Furthermore, Smad2 and Smad4 (Fig. 1a) or their isolated C domains (data not shown) interacted with the other's isolated C domain. In contrast to its dependence on TGF- β stimulation in mammalian cells³, the Smad2-Smad4 interaction occurred spontaneously in yeast, perhaps as a result of Smad phosphorylation by yeast kinases. These interactions were also analysed in mammalian cells by transfection of Smad2 and Smad4 fragments tagged with a Flag or HA epitope. The isolated C domains of Smad2 and 4 each

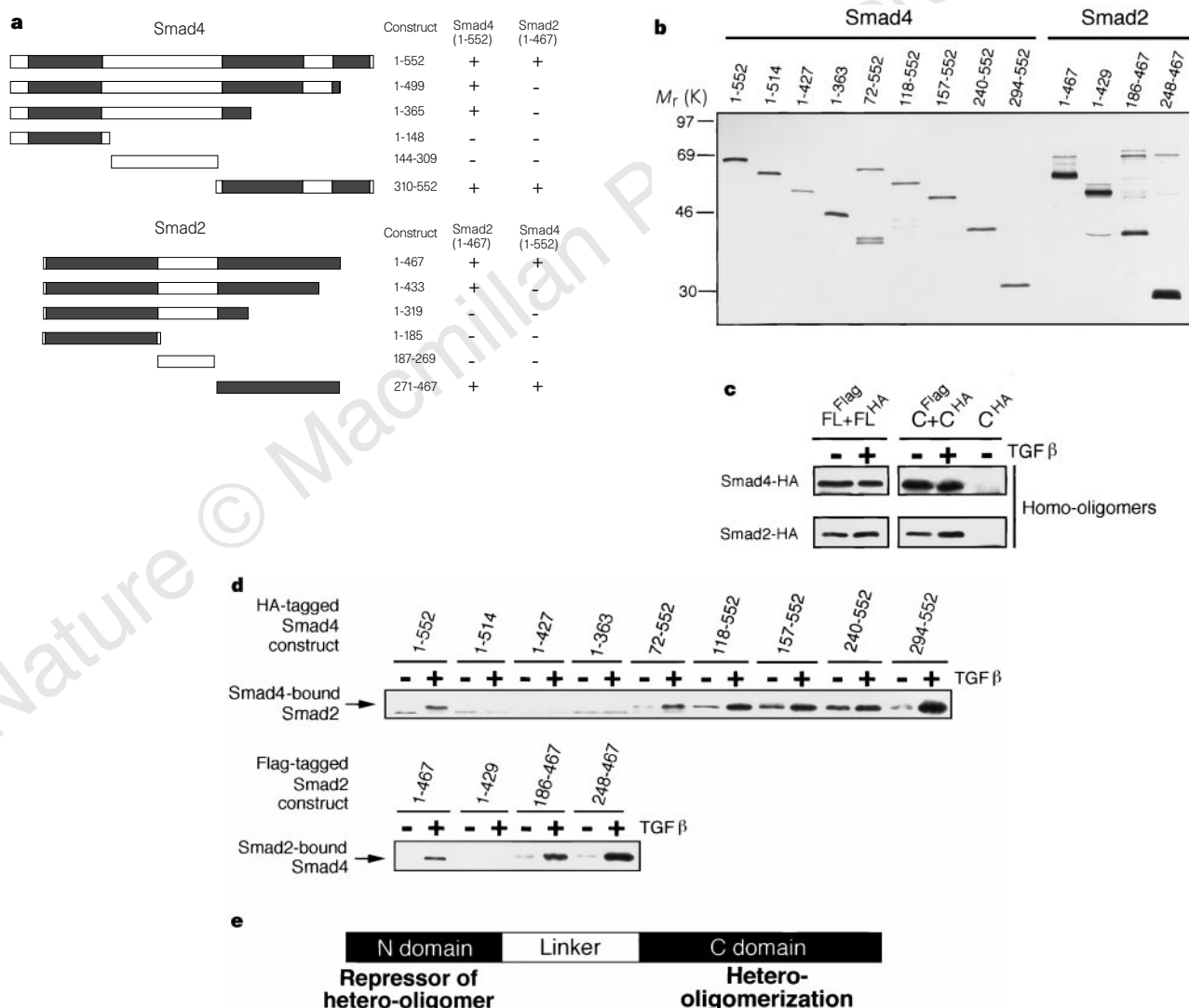


Figure 1 Analysis of Smad4 and Smad2 domain interactions. **a**, Smad4 and Smad2 interactions with themselves and each other in a yeast two-hybrid system. The indicated portions of Smad4 or Smad2 were tested for interaction with full-length or C domains of Smad2 or Smad4 fused to the LexA DNA-binding domain. **b**, Expression of HA-tagged Smad4 constructs and Flag-tagged Smad2 constructs was determined by epitope-tag immunoprecipitation from ³⁵S-methionine-labelled cells. **c**, Homo-oligomerization of Smad4 or Smad2 C domains. COS cells were transiently transfected with full-length (FL) Smad4 or Smad2 or their C domains (C) (Smad4 amino acids 294-552; Smad2 amino acids

248-467). Versions of the same protein tagged at the amino terminus with the Flag epitope or at the carboxy terminus with the HA epitope were co-transfected. Some cultures were incubated with TGF- β for 1 h before lysis. Homo-oligomerization was analysed by anti-HA immunoblotting of anti-Flag immunoprecipitates. **d**, Hetero-oligomerization of cotransfected HA-tagged Smad4 and Flag-tagged Smad2 deletion constructs. Smad2-Smad4 interaction was analysed by anti-Flag immunoblotting of anti-HA immunoprecipitates (top) or anti-HA immunoblotting of anti-Flag immunoprecipitates (bottom). **e**, Summary of SMAD domain contributions to Smad2-Smad4 hetero-oligomerization.

formed homo-oligomers in COS cells, as determined by co-immunoprecipitation of differently tagged constructs (Fig. 1c). In agreement with previous reports³, the Smad2–Smad4 interaction in COS cells requires TGF- β receptor stimulation (Fig. 1d), whereas the homo-oligomeric interactions do not (Fig. 1c). The C domains of Smad2 and Smad4 were necessary and sufficient for this interaction (Fig. 1d).

Deletion of the N domain in either Smad2 or Smad4 caused a constitutive association with the full-length version of the other protein, and this association was further enhanced by TGF- β action (Fig. 1d). The amount of expression of the Smad2 and Smad4 deletion products was similar to that of the full-length proteins (Fig. 1b), arguing that the constitutive association was not due to higher

expression of the N domain-deletion products. Thus, the combined results in yeast and COS cells indicate that the Smad2–Smad4 interaction is mediated primarily by the C domains, requires the integrity of these domains, and is inhibited by the presence of the N domains (Fig. 1e).

Consistent with this inhibitory function, expression of the Smad4 N domain (Fig. 2a) or the Smad2 N domain (Fig. 2b) inhibited the association between full-length as well as the C-domain forms of Smad2 and Smad4 in COS cells. This effect is specific because the overexpression of N domains does not inhibit homo-oligomerization of the C domains (Fig. 2c) or the expression of cotransfected C domains (data not shown). In yeast, the isolated N domains of Smad2 and 4 interacted weakly with their respective C domains (data not shown) and not at all with the full-length proteins (Fig. 1a). However, a fusion protein of glutathione-S-transferase (GST) with the Smad2 N domain specifically bound the Smad2 C domain when tested as recombinant proteins *in vitro* (Fig. 2d). Furthermore, the Smad2 and 4 N domains interacted with their corresponding C domains in COS cells (Fig. 5c). The more efficient interaction of the Smad N and C domains in COS cells may require a post-translational modification or an accessory factor.

As the formation of Smad2–Smad4 complexes requires Smad2 phosphorylation¹⁰ at C-terminal serines^{9,10}, we investigated how this

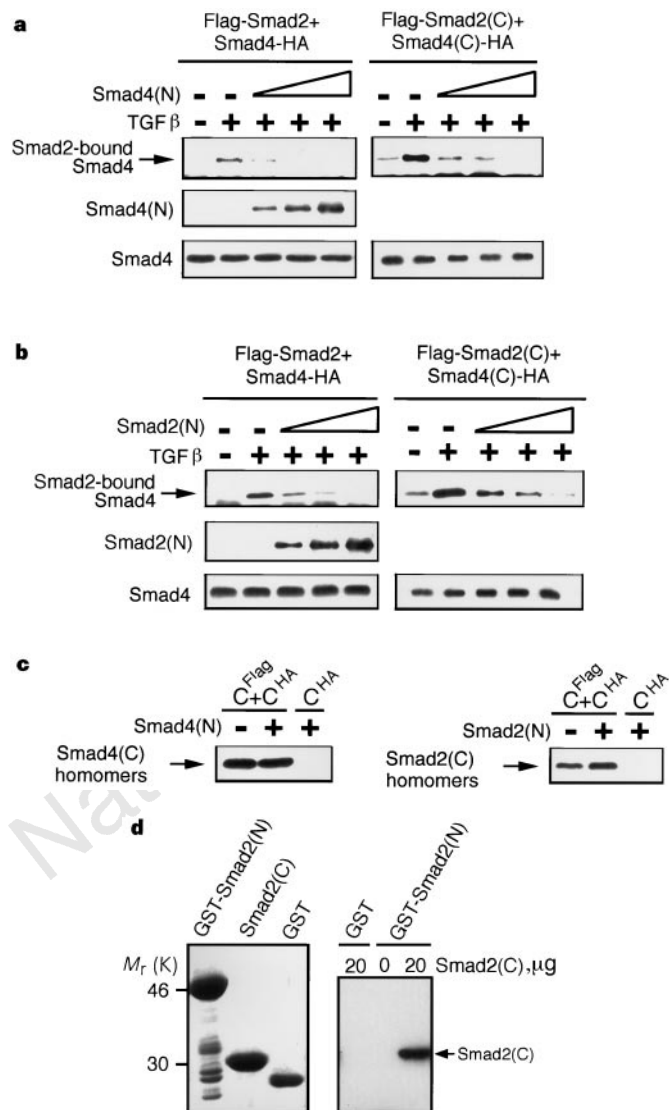


Figure 2 **a, b**, Inhibition of Smad2–Smad4 interaction by N domains. Increasing amounts (1, 2 and 4 μ g) of plasmid encoding the Smad4 N domain (amino acids 1–154) or the Smad2 N domain (amino acids 1–185) were co-transfected with the indicated full-length or C domain forms of Smad4 and Smad2. Smad2–Smad4 association was then determined. N domain and Smad4 expression was monitored by immunoblotting. **c**, N domain expression does not affect C domain homo-oligomerization. Flag-tagged and HA-tagged versions of Smad C domains were co-transfected with the indicated N domain. **d**, Specific interaction of the Smad2 N domain with the Smad2 C domain *in vitro*. Bacterially expressed proteins (Coomassie-blue-stained gel, left) were used to demonstrate Smad2(C) binding to GST–Smad2(N) but not to GST alone (right).

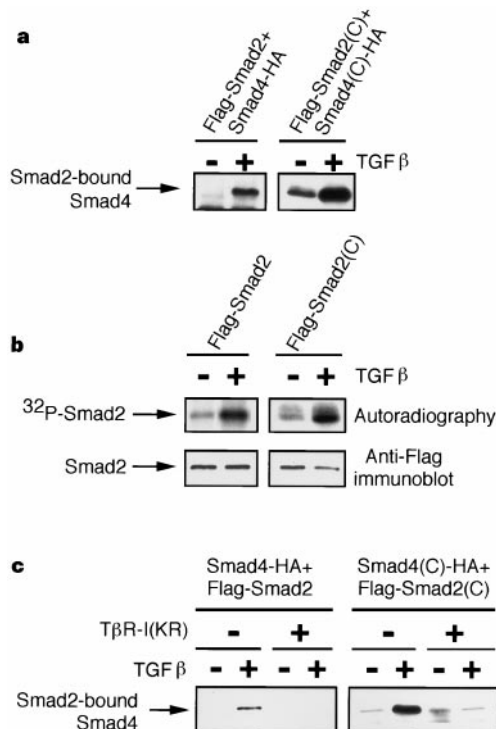


Figure 3 Effect of N domain deletion and agonist-induced phosphorylation on Smad2–Smad4 interaction. **a**, Constitutive association of the isolated C domains of Smad4 and Smad2, and further stimulation by TGF- β . The indicated full-length or C domain forms of Smad2 or Smad4 were cotransfected into COS cells. Cultures were stimulated with TGF- β , and Smad2–Smad4 interactions were analysed. **b**, Smad2 C domain phosphorylation in response to TGF- β . Constructs were transiently co-transfected with T β R-I into R-1B/L17 cells. Transfectants were labelled with ³²P-orthophosphate, stimulated with TGF- β for 20 min, and immunoprecipitated with anti-Flag antibody. **c**, The constitutive interaction of Smad4 and Smad2 C domains is independent of TGF- β -receptor-mediated phosphorylation. Smad2–Smad4 (full-length or C domain) complex formation was analysed in the presence or absence of a cotransfected dominant negative T β R-I construct (T β R-I(KR)).

process related to the inhibitory effect of the N domain. The Smad2 and Smad4 C domains spontaneously associated with each other and this association was further stimulated by TGF- β (Fig. 3a). The phosphorylation of the Smad2 C domain in response to TGF- β was comparable to that of full-length Smad2 (Fig. 3b). Co-transfection of a kinase-inactive, dominant-negative TGF- β type-I receptor [T β R-I(K232R)]¹³ abolished the TGF- β -stimulated association of Smad2 and Smad4 C domains but not their constitutive association (Fig. 3c). Thus the Smad2–Smad4 interaction is stimulated not only by phosphorylation of the Smad2 C domain¹⁰ but also by removal of the N domains.

We next investigated Smad2 and Smad4 products containing N-domain mutations that had been identified in human cancers^{2,14}. Smad4 mutations have been found in cancers of the pancreas, colon, oesophagus, breast, ovary, and head and neck^{1,14–17}; Smad2 mutations have been found in colon and head and neck cancers^{2,18,19}. Most of the missense mutations reported so far are in the C domains of Smad2 or 4. The three-dimensional structure of the Smad4 C domain predicts that some of these mutations destabilize the core

structure, others disrupt the C-domain homotrimer interface, and others disrupt a putative Smad4–Smad2 interface¹². Missense mutations have also been identified in the N domains of Smad2 and Smad4, including a Smad2 Arg133Cys mutation in a colon carcinoma² and a Smad4 Arg100Thr mutation in a pancreatic carcinoma¹⁴. Arg 133 in Smad2 corresponds to Arg100 in Smad4 (ref. 2), both of which are located in a highly conserved region of the SMAD proteins, suggesting that mutations at this residue are selected in cancer.

We confirmed that these N-domain mutations inactivate the signalling function of Smad2 and Smad4. In the *Xenopus* embryo, injection of Smad2 transcripts mimics the ability of activin to induce dorsal mesoderm in ectodermal explants^{5,20}. The Arg133Cys mutation prevented Smad2 from inducing muscle actin, a paraxial mesoderm marker, in this assay (Fig. 4a). In the human breast carcinoma cell line MDA-MB468, which has a Smad4 homozygous deletion¹⁴ and is therefore insensitive to TGF- β (ref. 3), Smad4 transfection restores TGF- β sensitivity; this sensitivity is enhanced by co-transfection of Smad2 (ref. 3). Using the TGF- β reporter gene construct 3TP-luciferase¹³, we found that mutation of Arg 133 in Smad2 or Arg100 in Smad4 stopped the co-transfected constructs from restoring TGF- β responsiveness in these cells (Fig. 4b). Moreover, Smad4 overexpression inhibits the proliferation of MDA-MB468 cells³, which was also disrupted by the Arg100Thr mutation (Fig. 4c).

To determine the basis for the inactivating effect of these N domain mutations, we transfected mutant Smad2 and Smad4 into COS cells and investigated their expression and interactions. Expression of both mutants was similar to wild type (Fig. 5a), as was their ability to form homo-oligomers (Fig. 5b). But both mutants failed to form Smad2–Smad4 complexes in response to TGF- β (Fig. 5b). As the Smad2–Smad4 interaction is primarily a function of the C domains and is repressed by the N domains, we investigated the inhibitory function of the wild-type and mutant N domains.

When expressed as separate polypeptides in COS cells, the N domains of Smad2 or Smad4 associate with the corresponding C domain (Fig. 5c). This association is not observed between the N domain of one SMAD and the C domain of the other (Fig. 5c). The Smad2 and Smad4 mutant N domains interacted with the corresponding C domains more strongly (18- and 22-fold, respectively) than did the wild-type N domains (Fig. 5c). As these mutations do not increase expression of the N domains, this increase in binding probably results from an increase in affinity of the mutant N domains for the C domains. Furthermore, the mutant N domains were more effective than wild-type N domains at inhibiting Smad2–Smad4 hetero-oligomerization (Fig. 5d).

To determine the effect of the Smad4 N domain on signalling by Smad2–Smad4, we tested its effect on the activation by Smad2 and Smad4 of the 3TP-luciferase reporter. In agreement with previous studies³, overexpression of Smad2 and Smad4 or of their C domains activated this reporter (Fig. 5e). Co-transfection of the Smad2 or Smad4 N domains significantly inhibited this effect (Fig. 5e). Furthermore, the mutant N domains were more potent inhibitors than the wild-type N domains of the Smad2/Smad4-mediated response (Fig. 5e). Thus, the binding of a Smad N domain to the corresponding C domain correlates with inhibition of the Smad2–Smad4 interaction and signalling function. We conclude that the Arg133Cys and Arg100Thr mutations increase the inhibitory function of the Smad2 and Smad4 N domains and inactivate signalling by Smad2–Smad4. As Smad4 partners other SMAD proteins besides Smad2 (refs 3, 7), the Arg100Thr mutation should disrupt signalling by these proteins as well.

We have shown that the N domain in SMAD proteins interacts with and represses the effector function of the C domain, and that certain Smad2 and Smad4 mutations found in human cancer cells inactivate these proteins by augmenting the inhibitory function of

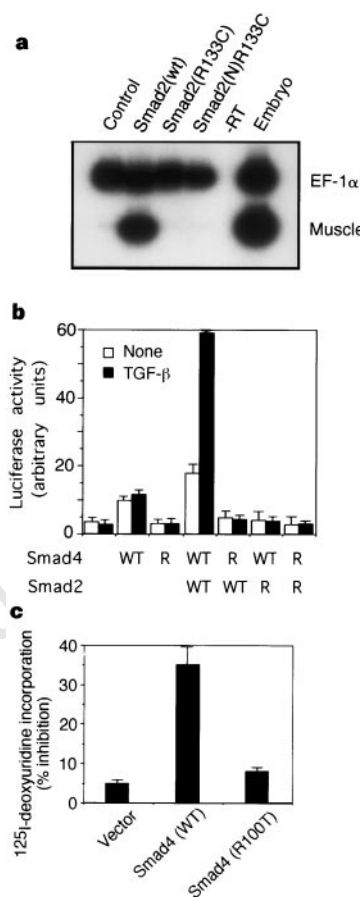


Figure 4 Biological activity of Smad2 and Smad4 containing tumour derived N-domain mutations. **a**, Wild-type Smad2 induces the paraxial mesoderm marker muscle actin in *Xenopus* ectodermal explants, whereas Smad2(R133C) or its N domain alone (Smad2(N/R133C)) are unable to induce it. EF-1 α was used as an internal control. **b**, Cotransfection of wild-type Smad2 and Smad4 (WT) restores TGF- β responsiveness into Smad4-defective MDA-MB468 breast cancer cells, whereas cotransfections including the Smad2(R133C) mutant (R), the Smad4(R100T) mutant (R), or both mutants, do not. The TGF- β responsiveness of these cells was determined using the reporter construct 3TP-lux. **c**, Overexpression of wild-type Smad4 inhibits MDA-MB468 cell proliferation, whereas overexpression of the Smad4(R100T) mutant does not. The proliferation of the cells was determined by measuring iododeoxyuridine incorporation into DNA. Results are the average $\pm$ s.d. of triplicate assays.

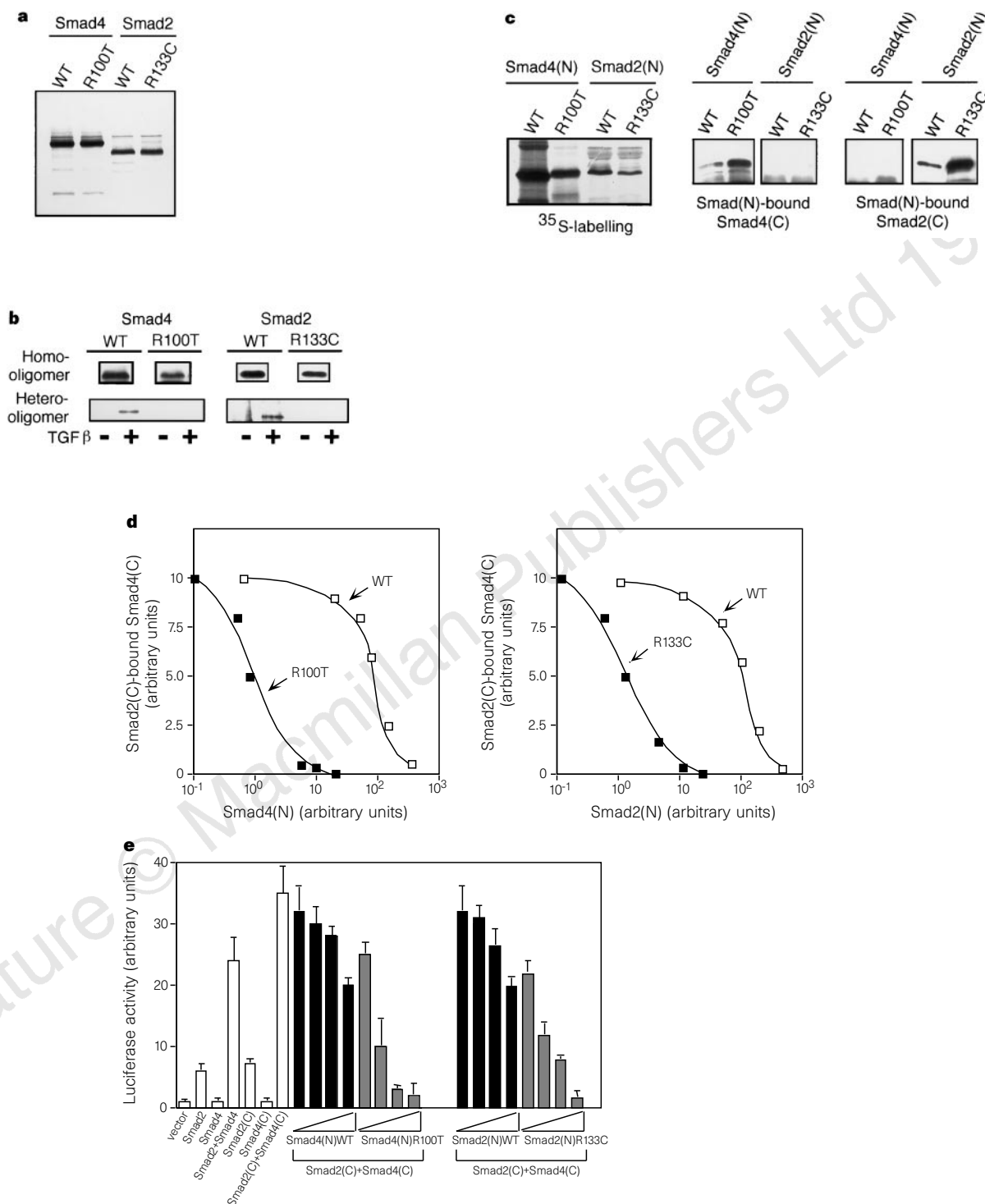


Figure 5 Gain of autoinhibitory function of Smad4 and Smad2 N domain mutants. **a**, Expression of wild-type and mutant Smads were determined by epitope-tag immunoprecipitation from ³⁵S-methionine-labelled transfected COS cells. **b**, N domain mutations inhibit Smad2-Smad4 interaction. HA-tagged wild-type or mutant Smad4 was co-transfected with Flag-tagged Smad4 (for homo-oligomeric interaction) or Flag-tagged Smad2 (for hetero-oligomeric interaction) in COS cells. Likewise, Flag-tagged wild-type or mutant (R133C) Smad2 was cotransfected with HA-tagged Smad2 or HA-tagged Smad4. **c**, N domain interaction with the C domain, and its enhancement by mutation. Flag-tagged N domains were cotransfected with the HA-tagged C domains indicated at the bottom. N domain interaction with the C domain was determined by anti-HA immunoblotting of anti-Flag immunoprecipitates. N domain expression was monitored by immunoprecipitation from ³⁵S-methionine-labelled cells. **d**, Mutant N domains

inhibit the Smad2-Smad4 interaction more strongly. Increasing amounts of plasmid DNA encoding wild-type or mutant (R100T) Smad4 N domain (left) or wild-type or mutant (R133C) Smad2 N domain (right) were cotransfected with Flag-tagged Smad2 C domain and HA-tagged Smad4 C domain. The amounts of Smad4 or Smad2 N domain protein and Smad2(C)-bound Smad4(C) were determined by immunoblotting, quantified (ImageQuant; Molecular Dynamics) and plotted against each other. **e**, N domain inhibition of Smad2-Smad4 signalling. R-1B/L17 cells were transiently transfected with the indicated constructs and 3TP-lux reporter. Amounts of transfected Smad4 and Smad2 were adjusted³ to increase luciferase expression synergistically. Increasing amounts (1, 2, 4 and 6 µg) of plasmid DNA encoding wild-type or mutant N domains were cotransfected with a Smad4(C)/Smad2(C) combination. Results (luciferase activity in arbitrary units) are the average ± s.d. of triplicate assays.

the N domain. Other tumour-suppressor mutations, including missense mutations in the C domains of Smad2 and Smad4, act by disrupting protein stability or effector function¹². Our findings reveal a mechanism of tumour suppressor inactivation which instead involves a gain of autoinhibitory function. Antagonists of SMAD autoinhibition might be useful in reversing the effects of this type of mutation. □

Methods

Construction of expression vectors. To generate human Smad4 and Smad2 mutations, a fragment of the corresponding cDNAs^{1,4} was amplified by PCR. The amplified region was subcloned into full-length Smad4 or Smad2 in pCMV5 for transfection into mammalian cells. The regions amplified by PCR and the presence of missense mutations were confirmed by sequencing.

Yeast two-hybrid system. LexA fusions were created in pBTM 116 (ref. 21) and GAD fusions within pGAD424 (Clontech). Interactions were tested in the strain L40 (ref. 22). Activation of the LexA operator–HIS3 reporter was assayed on medium lacking histidine with increasing concentrations of 3-amino-triazole.

Transfection, immunoprecipitation, immunoblot and metabolic labelling. For Smad2/Smad4 homo- or hetero-complex analysis, COS cells were transiently transfected with the indicated constructs and stimulated with 200 pM TGF- β 1 for 1 h. Cells were lysed in TNE buffer³, immunoprecipitated with anti-Flag M2 monoclonal antibody (IBI; Eastman Kodak), and interacting proteins detected by immunoblotting with the anti-HA monoclonal antibody 12CA5 (Boehringer Mannheim) as described³. Anti-SMAD rabbit polyclonal antibody was raised against full-length Smad1. To study the interaction between the N and C domains of Smad4 or Smad2, transiently transfected COS cells were lysed in LSLD buffer (50 mM HEPES, pH 7.4, 50 mM NaCl, 0.1% Tween 20, 10% glycerol, 1 mM DTT) containing protease and phosphatase inhibitors. Immunoprecipitation and immunoblotting were done as described. COS or R-1B/L17 cells transfected with the indicated constructs were labelled with ³⁵S-methionine or ³²P-orthophosphate and visualized by electrophoresis and autoradiography²³.

In vitro binding. The Smad2 N domain (amino acids 1–184) fused to GST and the Smad2 C domain (amino acids 248–467) protein were expressed in *Escherichia coli* and partially purified by column chromatography. GST fusion proteins were conjugated to glutathione–Sepharose beads (Pharmacia) and incubated with Smad2 C domain protein. Smad2 C domain protein was detected by western blotting using anti-Smad antibody.

Functional assays. For the animal cap assay, RNA (10 nl, 2 ng) was introduced in the animal pole of two-cell *Xenopus* embryos. Animal caps were explanted at the blastula stage and cultured to the tadpole stage. Total RNA from the explants and control sibling embryos was extracted and was amplified with reverse transcription by using muscle actin and *EF-1 α* primers²⁴. In the MDA-MB468 cell experiments, the amount of transfected plasmids was adjusted to render the TGF- β response dependent on both Smad2 and Smad4. Luciferase and growth inhibition were assayed as described³.

Received 24 March; accepted 6 May 1997.

- Hahn, S. A. *et al.* DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science* **271**, 350–353 (1996).
- Eppert, K. *et al.* MADR2 maps to 18q21 and encodes a TGF- β -regulated MAD-related protein that is functionally mutated in colorectal carcinoma. *Cell* **86**, 543–552 (1996).
- Lagna, G., Hata, A., Hemmati-Brivanlou, A. & Massagué, J. Partnership between DPC4 and SMAD proteins in TGF- β signalling pathways. *Nature* **383**, 832–836 (1996).
- Liu, F. *et al.* A human Mad protein acting as a BMP-regulated transcriptional activator. *Nature* **381**, 620–623 (1996).
- Baker, J. & Harland, R. M. A novel mesoderm inducer, mMad2-2, functions in the activin signal transduction pathway. *Genes Dev.* **10**, 1880–1889 (1996).
- Massagué, J., Hata, A. & Liu, F. TGF- β signalling through the Smad pathway. *Trends Cell Biol.* **7**, 187–192 (1997).
- Zhang, Y., Feng, X.-H., Wu, R.-Y. & Derynck, R. Receptor-associated Mad homologues synergize as effectors of the TGF- β response. *Nature* **383**, 168–172 (1996).
- Hoodless, P. A. *et al.* MADR1, a MAD-related protein that functions in BMP2 signaling pathways. *Cell* **85**, 489–500 (1996).
- Macías-Silva, M. *et al.* MADR2 is a substrate of the TGF β receptor and phosphorylation is required for nuclear accumulation and signaling. *Cell* **87**, 1215–1224 (1996).
- Kretschmar, M., Liu, F., Hata, A., Doody, J. & Massagué, J. The TGF- β family mediator Smad1 is phosphorylated directly and activated functionally by the BMP receptor kinase. *Genes Dev.* **11**, 984–995 (1997).
- Wu, R.-Y., Zhang, Y., Feng, X.-H. & Derynck, R. Heteromeric and homomeric interactions correlated with signaling activity and functional cooperativity of Smad3 and Smad4/DPC4. *Mol. Cell. Biol.* **17**, 2521–2528 (1997).

- Shi, Y., Hata, A., Lo, R. S., Massagué, J. & Pavletich, N. P. A structural basis for mutational inactivation of the tumour suppressor Smad4. *Nature* **388**, 87–93 (1997).
- Cárcamo, J. *et al.* Type I receptors specify growth inhibitory and transcriptional responses to TGF- β and activin. *Mol. Cell. Biol.* **14**, 3810–3821 (1994).
- Schutte, M. *et al.* DPC4 gene in various tumor types. *Cancer Res.* **56**, 2527–2530 (1996).
- Kim, S. K. *et al.* DPC4, a candidate tumor suppressor gene, is altered infrequently in head and neck squamous cell carcinoma. *Cancer Res.* **56**, 2519–2521 (1996).
- Nagatake, M. *et al.* Somatic *in vivo* alterations of the DPC4 gene at 18q21 in human lung cancers. *Cancer Res.* **56**, 2718–2720 (1996).
- Barrett, M. T., Schutte, M., Kern, S. E. & Reid, B. J. Allelic loss and mutational analysis of the DPC4 gene in esophageal adenocarcinoma. *Cancer Res.* **56**, 4351–4353 (1996).
- Riggins, G. J. *et al.* Mad-related genes in the human. *Nature Genet.* **13**, 347–349 (1996).
- Uchida, K. *et al.* Somatic *in vivo* alterations of the JV18-1 gene at 18q21 in human lung cancers. *Cancer Res.* **56**, 5583–5585 (1996).
- Graff, J. M., Bansal, A. & Melton, D. A. *Xenopus* Mad proteins transduce distinct subsets of signals for the TGF β superfamily. *Cell* **85**, 479–487 (1996).
- Bartel, P. L., Chien, C.-T., Sternglanz, R. & Fields, S. in *Cellular Interactions in Development: A Practical Approach* (ed. Hartley, D. A.) 153 (Oxford University Press, Oxford, 1993).
- Vojtek, A. B., Hollenberg, S. M. & Cooper, J. A. Mammalian Ras interacts directly with the serine/threonine kinase Raf. *Cell* **74**, 205–214 (1993).
- Wrana, J. L., Attisano, L., Wieser, R., Ventura, F. & Massagué, J. Mechanism of activation of the TGF- β receptor. *Nature* **370**, 341–347 (1994).
- Hemmati-Brivanlou, A. & Melton, D. A. Inhibition of activin receptor signalling promotes neuralization in *Xenopus*. *Cell* **77**, 273–281 (1994).

Acknowledgements. We thank Y. Shi for recombinant Smad2 proteins, and I. Reynisdóttir, J. Doody and S. Lee for advice and technical assistance. G.L. thanks A. Hemmati-Brivanlou for support and advice. This work was supported by NIH Breast Spore and Cancer Center grants. D.W. is the recipient of a postdoctoral fellowship from the Human Frontier Science Program. A.H. is a research associate and J.M. an investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to J.M. (e-mail: j-massague@ski.mskcc.org).

A structural basis for mutational inactivation of the tumour suppressor Smad4

Yigong Shi, Akiko Hata*, Roger S. Lo*, Joan Massagué* & Nikola P. Pavletich

Cellular Biochemistry and Biophysics Program, *Cell Biology Program and the Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA

The Smad4/DPC4 tumour suppressor¹ is inactivated in nearly half of pancreatic carcinomas² and to a lesser extent in a variety of other cancers^{3–4}. Smad4/DPC4, and the related tumour suppressor Smad2, belong to the SMAD family of proteins that mediate signalling by the TGF- β /activin/BMP-2/4 cytokine superfamily from receptor Ser/Thr protein kinases at the cell surface to the nucleus^{5–7}. SMAD proteins, which are phosphorylated by the activated receptor, propagate the signal, in part, through homo- and hetero-oligomeric interactions^{8–13}. Smad4/DPC4 plays a central role as it is the shared hetero-oligomerization partner of the other SMADs. The conserved carboxy-terminal domains of SMADs are sufficient for inducing most of the ligand-specific effects, and are the primary targets of tumorigenic inactivation. We now describe the crystal structure of the C-terminal domain (CTD) of the Smad4/DPC4 tumour suppressor, determined at 2.5 Å resolution. The structure reveals that the Smad4/DPC4 CTD forms a crystallographic trimer through a conserved protein–protein interface, to which the majority of the tumour-derived missense mutations map. These mutations disrupt homo-oligomerization *in vitro* and *in vivo*, indicating that the trimeric assembly of the Smad4/DPC4 CTD is critical for signalling and is disrupted by tumorigenic mutations.

The conserved C-terminal domain can mediate many of the biological effects of SMAD proteins, and the conserved N-terminal domain can negatively regulate the SMAD activity^{14,15}. When over-expressed in a Smad4/DPC4^{–/–} cell line, the Smad4/DPC4 CTD can activate the transcription of TGF- β responsive genes and result in



Figure 1 The structure of the Smad4/DPC4 CTD consists of a β -sandwich with a three-helix bundle on one end and a collection of three large loops and an α -helix on the other. The view is along the edge of the β -sandwich; the dotted line represents the disordered region between the H3 and H4 helices. Figures were prepared with the programs MOLSCRIPT²⁶ and RASTER3D²⁷.

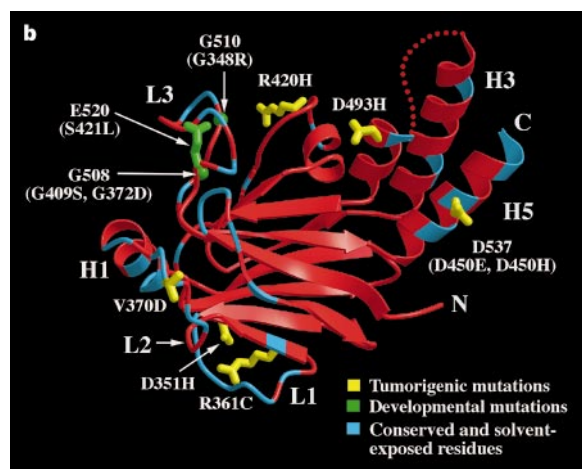
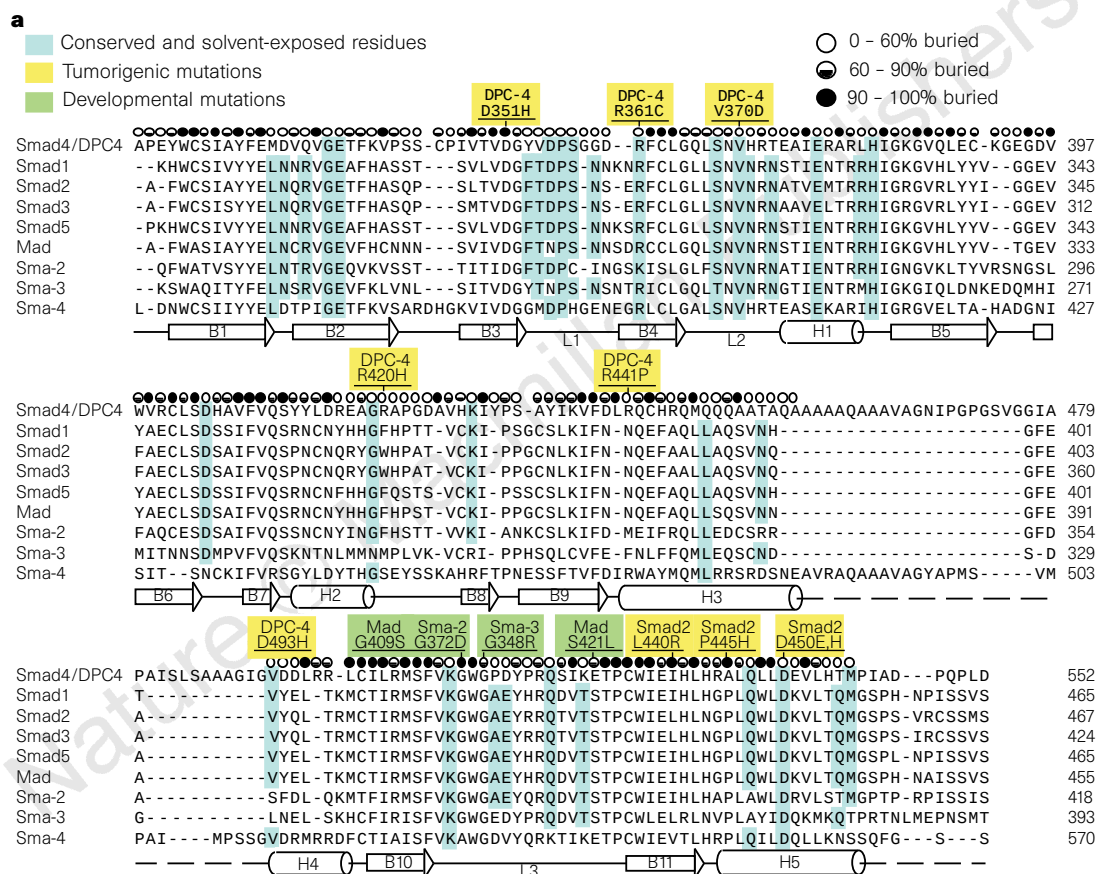


Figure 2 SMAD C-terminal domains are highly conserved and are targeted by tumorigenic and developmental mutations. **a**, Sequence alignment of C-terminal domains of five human SMAD proteins^{18,19} (Smad 1, 2, 3, 5 and Smad4/DPC4) and homologues from *Drosophila*¹⁸ (Mad) and *C. elegans*¹⁹ (Sma-2, 3, 4), with the Smad4/DPC4 CTD secondary structure elements indicated below the sequences. Residues that are more than 40% exposed to solvent have no significant role in the structure and are conserved in at least 6 out of the 9 aligned sequences are highlighted in cyan. The 14 missense mutations shown above the alignment include tumour-derived Smad4/DPC4 and Smad2 mutations^{12,41,27,28} (yellow), as well as mutations from *Drosophila* and *C. elegans* genetic screens^{18,19} (developmental mutations, in green). The residues at which these mutations occur are in bold face and underlined. **b**, Mapping of the missense mutations and the highly conserved and solvent-exposed residues identifies the three-helix bundle and the three-loop/helix region as regions likely to be important for the macromolecular recognition that mediates SMAD function. Colour coding is the same as in **a**. The amino-acid substitution and the residue number of the mutation in SMAD family members other than Smad4/DPC4 are shown in parentheses. The three structural mutations (Arg441Pro in Smad4/DPC4, Leu440Arg and Pro445H in Smad2) are not shown.

Table 1 Statistics from the crystallographic analysis

Data set		Native 1 (8 °C)	Native 2 (−170 °C)	Thimerosal	UO ₂ (OAc) ₂ *	(CH ₃) ₃ PbOAc*		
Resolution (Å)		3.0	2.5	3.0	3.0	3.2		
Observations		30,691	39,125	30,572	23,488	25,150		
Unique reflections		7,189	11,496	7,073	6,765	5,759		
Data coverage (%)		96.8	96.5	96.8	92.8	94.6		
<i>R</i> _{sym} (%)†		6.5	3.7	4.8	8.5	10.0		
MIR analysis (20.0–3.2 Å):								
Mean isomorphous difference‡				0.18	0.14	0.24		
Phasing power§				2.54	1.38	1.03		
Refinement statistics:					R.m.s.d.#			
Resolution (Å)	Reflections (<i>F</i> > 2σ)	Protein atoms	Waters atoms	<i>R</i> -factor (%)	<i>R</i> -free¶ (%)	Bonds (Å)	Angles (°)	<i>B</i> -factor (Å ²)
7.0–2.5	10,359	1,522	129	20.9	28.6	0.010	1.66	3.29

* OAc, acetate.

† $R_{\text{sym}} = \sum_i \sum_h |I_{h,i} - \bar{I}_h| / \sum_i \sum_h I_{h,i}$ for the intensity (I) of i observations of reflection h .

‡ Mean isomorphous difference = $\sum |F_{\text{PH}} - F_{\text{P}}| / \sum F_{\text{PH}}$, where F_{PH} and F_{P} are the derivative and native structure factors, respectively.

§ Phasing power = $[(F_{\text{calc}})^2 / (F_{\text{PH(obs)}} - F_{\text{PH(calc)}})^2]^{1/2}$.

|| R -factor = $\sum |F_{\text{obs}} - F_{\text{calc}}| / \sum |F_{\text{obs}}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.

¶ R -free is the R -factor calculated using 5% of the reflection data chosen randomly and omitted from the start of refinement.

R.m.s.d., root-mean-square deviations from ideal geometry and root-mean-square variation in the B -factor of bonded atoms.

growth arrest in a ligand-independent manner, paralleling the effects of the TGF- β ligand⁹; microinjection of messenger RNAs encoding the CTD of Smad2 into *Xenopus* embryos can induce a mesoderm response that mimics the effects of the full-length protein¹⁶; and the Smad4/DPC4 CTD fused to a heterologous DNA-binding domain can activate gene expression from a reporter construct¹⁴. Consistent with the SMAD C-terminal domain being the main effector domain, the majority (10 out of 13) of tumorigenic missense mutations in Smad4/DPC4 and Smad2, as well as mutations isolated from *Drosophila* and *Caenorhabditis elegans* genetic screens map to the C-terminal domain. To investigate how the SMAD C-terminal domain functions in mediating TGF- β signalling and how its mutation in cancer inactivates the pathway, we determined the crystal structure of the 234-amino-acid Smad4/DPC4 CTD (residues 319–552) at 2.5 Å resolution (Table 1).

The structure consists of a β -sandwich with twisted antiparallel β -sheets of five and six strands each (Fig. 1). One end of the β -sandwich is capped by a three- α -helix bundle (H3, H4 and H5 helices) that extends over the plane of the six-stranded β -sheet, at a roughly perpendicular angle; the other end of the β -sandwich is capped by a group of three large loops and an α -helix (L1, L2, L3 loops and H1 helix; Fig. 1). To simplify the presentation, the three large loops and α -helix, as well as portions of β -strands in their immediate vicinity, will be collectively referred to as the loop/helix region. The three α -helices of the bundle pack in an up-down-up orientation primarily through leucine residues. In between the H3 and H4 helices, a 34-amino-acid sequence that is rich in Ala (39%), Gly and Pro residues and is present only in Smad4/DPC4 and its *C. elegans* homologue Sma-4, is disordered in the crystals (residues 457–491). In the loop/helix region, the L1, L2 and L3 loops of 7, 9 and 18 residues, respectively, and the H1 helix are mostly polar and pack through extended hydrogen-bond networks. These hydrogen bonds are likely to contribute to the rigid structure of this region that is suggested by the well defined electron density.

SMAD proteins are highly conserved within the family and across species, with Smad4/DPC4 and its *C. elegans* homologue, Sma-4, representing a somewhat divergent subtype which still retains about 40% identity with other family members^{5,7} (Fig. 2a). Many of the conserved residues have structural roles. These include the hydrophobic residues that make up the hydrophobic core of the β -sandwich and of the three-helix bundle, as well as many of the polar residues that form the hydrogen-bond networks important for the structure of the loop/helix region. Examples of the latter group are the invariant Arg 372 and Arg 380 residues from the H1 helix that make four and three charge-stabilized hydrogen bonds,

respectively. Many other highly conserved residues are exposed to solvent and do not appear to stabilize the structure. They are thus candidates for functional residues that may mediate macromolecular interactions important for the function of SMAD proteins. The structure reveals that these candidate functional residues (highlighted in Fig. 2b) have a strong tendency to cluster at the loop/helix region and the three-helix bundle.

Another indication that the loop/helix region and the three-helix bundle are functionally important comes from an analysis of nine tumour-derived missense mutations, some occurring several times, in the CTDs of the Smad4/DPC4 and Smad2 tumour suppressors. Excluding three mutations that map to structural residues, five of the six tumour-derived missense mutations map to either the loop/helix region or to the three-helix bundle: the Smad4/DPC4 mutations Asp351His², Arg361Cys¹⁷ and Val370Asp¹⁷ map to the loop/helix region, whereas the Smad4/DPC4 mutation Asp493His¹ and the Smad2 mutation Asp450Glu¹² (corresponding to Asp 537 of Smad4/DPC4) map to the three-helix bundle. We presume that these mutations deprive the CTD of critical intermolecular contacts. The one mutation that does not map to either region is Arg420His from Smad4/DPC4, which instead maps to the side of the β -sandwich (H2 helix), a region that is not as well conserved. The remaining three mutations map to structural residues: the Smad2 Leu440Arg mutation (corresponding to Ile 527 of Smad4/DPC4) in the hydrophobic core of the β -sandwich probably disrupts packing in the hydrophobic core; the Smad4/DPC4 Arg441Pro mutation at the three-helix bundle probably disrupts the H3 helix because of the introduction of a proline in the middle of the helix; and the Smad2 Pro445His mutation (corresponding to Ala 532 in Smad4/DPC4), also at the three-helix bundle, probably disrupts the packing between the three-helix bundle and the β -sandwich as there is little space for the larger histidine side chain in this part of the hydrophobic core.

In support of the functional significance of the loop/helix region, mutations in *Drosophila* and *C. elegans* in this region produce null or severe developmental phenotypes^{18,19}. These mutations map to Gly 508 (*Drosophila* Mad, *C. elegans* Sma-2), Gly 510 (Sma-3), and Glu 520 (Mad) of the L3 loop in the loop/helix region (Fig. 2). Thus the location of conserved, solvent-exposed residues and of mutations derived from tumours or from *Drosophila* and *C. elegans* genetic screens together indicate that the loop/helix region and the three-helix bundle are critical for mediating Smad activity.

As SMAD CTDs can mediate most of the effects of the full-length proteins, we tested the Smad4/DPC4 CTD for homo-oligomerization activity. Co-immunoprecipitation using extracts from COS

cells transfected with differentially tagged Smad4/DPC4 CTD constructs showed that the Smad4/DPC4 CTD could still form homo-oligomers when overexpressed¹⁵ (Fig. 3d), suggesting that the CTD may contain a primary homo-oligomerization activity. However, full-length Smad4/DPC4 homo-oligomers are more stable than Smad4/DPC4 CTD homo-oligomers *in vivo*¹⁵, indicating that residues N-terminal to the Smad4/DPC4 CTD may contribute to homo-oligomerization.

To investigate the homo-oligomerization activity of the Smad4/DPC4 CTD further, we examined the crystal packing of the Smad4/DPC4 CTD molecules and identified a trimer that formed through three identical extended protein–protein interfaces, burying 4,800 Å² of surface area (Fig. 3a). Each interface forms through

the interactions of the highly conserved regions of the Smad4/DPC4 CTD that contain most of the candidate functional residues: the loop/helix region of one subunit packs extensively with the three-helix bundle from another subunit while making a few additional contacts to residues from the β-sandwich (Fig. 3a). The only portion of the loop/helix region that does not participate in this interface is the L3 loop.

The trimer interface includes the majority of the conserved residues and the tumour-derived non-structural missense mutations (five out of six). Most noteworthy is an extended intermolecular hydrogen-bond network involving, from one subunit, the Arg 361 and Asp 351 side chains and two backbone amide groups of the loop/helix region, and from another subunit, the Asp 537 side

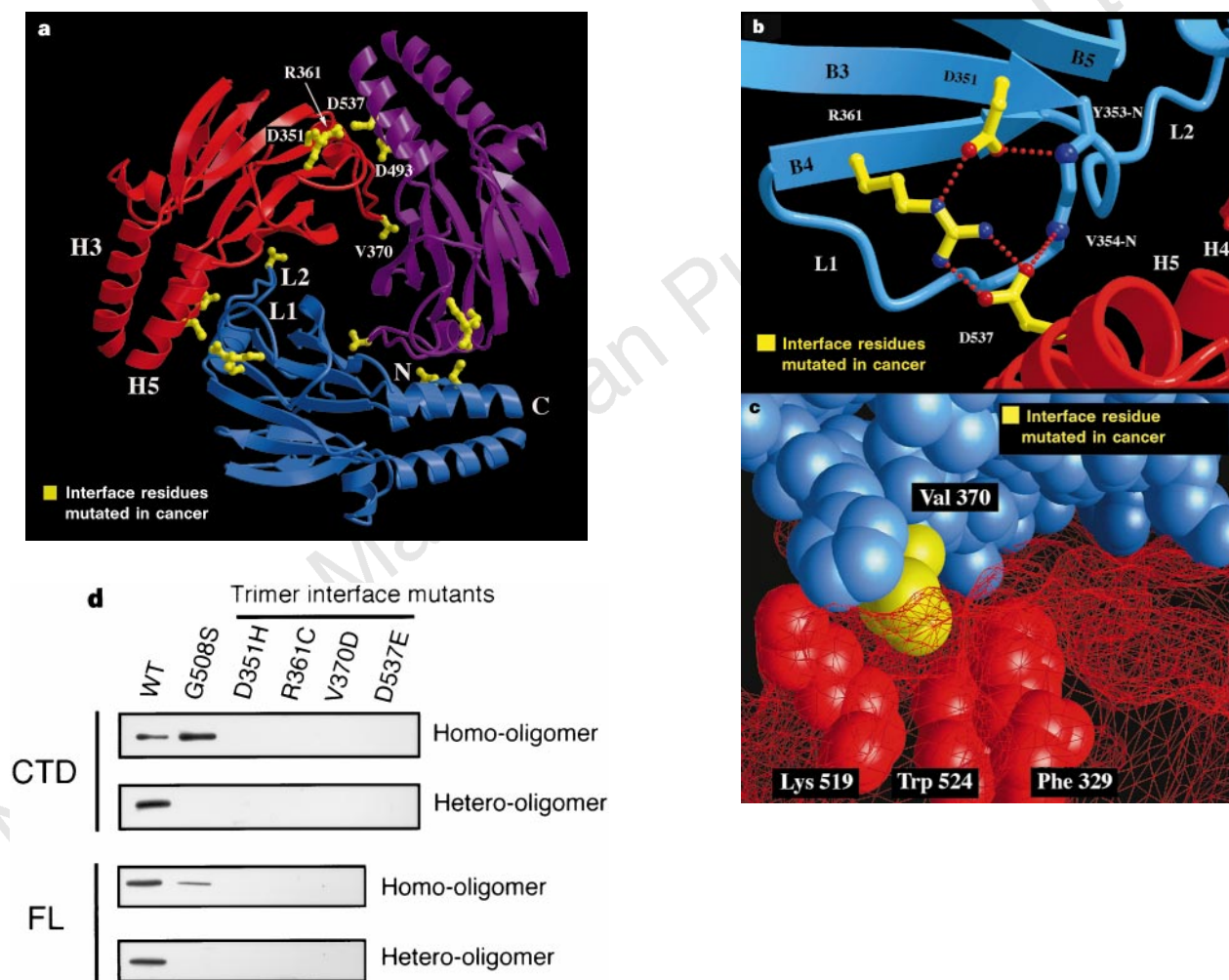


Figure 3 In the crystals, the Smad4/DPC4 CTD forms a trimer that is disrupted by tumorigenic mutations and is probably important for SMAD function. **a**, The three monomers (red, blue and purple) pack across three identical protein–protein interfaces. Tumour-derived missense mutations map to five amino acids (yellow) that are involved in intermolecular contacts. **b**, Close-up of an intermolecular hydrogen-bond network involving three residues, all of which have been found mutated in cancer. Colouring is as in **a**. **c**, Close-up showing the intermolecular packing of Val 370, mutated to Asp in colon tumours, against Phe 329, Trp 524 and the aliphatic portion of Lys 519. The subunit in which Val 370 is shown is in space-filling representation, whereas the other subunit is shown as a molecular surface (red mesh). Other intermolecular interactions not mentioned in the text include: van der Waals contacts between the L1 loop of the loop/helix region and the H4 and H5 helices of the three-helix bundle (Tyr 353, Val 354 and Pro 356 wedging in between His 530, Leu 533, Leu 536, Leu 540 and His 541); the hydrogen-bond networks between Ser 368 of the L2 loop and Arg 496, Glu 526 and His 528 of the

β-sheet, and between His 371 of the L2 loop and Asp 332 of the β-sheet. The figure was prepared with the program GRASP²⁹. **d**, *In vivo*, tumour-derived trimer interface mutations (D351H, R361C, V370D, D537E) disrupt both homo- and hetero-oligomerization, whereas a developmental mutation in the L3 loop (G508S) disrupts only hetero-oligomerization. To assay for homo-oligomerization, mammalian COS-1 cells were transiently transfected with Flag-tagged wild-type Smad4/DPC4 CTD and HA-tagged WT or mutant constructs. For hetero-oligomerization, cells were transfected with Flag-tagged Smad2 C-terminal domain and HA-tagged Smad4/DPC4 C-terminal domain WT or mutant constructs together with a constitutively active TGF-β type I receptor construct. The cell lysate was immunoprecipitated with anti-Flag antibody and subsequently immunoblotted using anti-HA antibody. Immunoblots indicated that the mutant Smad4/DPC4 CTDs were expressed at levels comparable to those of the wild-type constructs (data not shown). Lower panels, full-length (FL) proteins were treated similarly.

chain of the three-helix bundle (Fig. 3b). The Asp351, Arg 361, and Asp 537 residues are essentially invariant, apart from a conservative Arg-to-Lys substitution in Smad2 (Fig. 2a), and all three are mutated in cancer (Fig. 2): Asp351His and Arg361Cys mutations have been found in Smad4/DPC4 in ovarian² and colon cancers¹⁷, respectively, and the Asp450Glu mutation, corresponding to Asp 537 of Smad4/DPC4, has been isolated from Smad2 in colon cancer¹². These mutations all disrupt the intricate hydrogen-bond network at the interface. Also noteworthy are the intermolecular van der Waals contacts between Val 370 on the L2 loop of the loop/helix and Trp 524, Phe 329, and the aliphatic portion of the Lys 519 side chain on the β -sheet at the base of the three-helix bundle (Fig. 3c). The two aromatic residues are invariant, except for a conservative Tyr-to-Phe substitution in Smad4/DPC4 (Fig. 2a). Also, Val 370 is mutated to Asp in colon cancer¹⁷, causing the hydrophobic portion of the trimer interface to be destabilized by the charged amino acid. The Smad4/DPC4 Asp493His mutation found in pancreatic tumours¹ also maps to the trimer interface (Fig. 3a) and may interfere with the electrostatic packing of Asp 493 of one subunit with Arg 496 and Arg 497 of another subunit at the trimer interface. Asp 493 is near the disordered region of the H4 helix and its interaction with the arginine residues is not well defined.

Many of the other trimer-interface contacts are also conserved in the SMAD family (Fig. 3c legend), suggesting that other SMAD CTDs may form a similar trimeric structure. Not all residues in the Smad4/DPC4 CTD trimer interface are conserved in all SMAD, and those that differ may contribute to subtype specificity. An example is an intermolecular hydrogen-bond contact between His 371 and Asp 332: this pair is conserved in the *C. elegans* Smad4/DPC4 homologue Sma4, but is an invariant Asn-Asn pair in the pathway-restricted SMADs (Fig. 2).

If the trimeric Smad4/DPC4 CTD assembly in the crystals is part of the *in vivo* homo-oligomer, then mutations of residues that make intermolecular contacts at the interface, particularly tumour-derived mutations, should disrupt or reduce homo-oligomerization *in vivo*. Figure 3d shows the results of co-immunoprecipitation experiments using extracts from COS cells transfected with differentially tagged mutant Smad4/DPC4 molecules. All four of the tumorigenic mutations that are important at the trimer interface, Asp 351, Arg 361, Val 370 and Asp 537, disrupt homo-oligomerization of the Smad4/DPC4 CTD and of full-length Smad4/DPC4 (Fig. 3d). Conversely, the *Drosophila/C. elegans* developmental mutation Gly508Ser (Fig. 2a) had no effect on homo-oligomerization (Fig.

3d). This mutation maps to the L3 loop, which is the only portion of the loop/helix region not involved in the trimer interface.

If Smad4/DPC4 CTD forms a trimer, then so should full-length Smad4/DPC4. Recombinant full-length Smad4/DPC4, purified to near homogeneity, elutes from a gel-filtration column with an apparent relative molecular mass (M_r) of ~ 180 K, consistent with the calculated size (181K) of the Smad4/DPC4 trimer (Fig. 4a). This large M_r is probably due to trimerization, because the M_r is reduced by a factor of ~ 3 in the tumour-derived trimer-interface mutants (Fig. 4b). Conversely, the *Drosophila/C. elegans* developmental mutation Gly508Ser in the L3 loop has no effect on the large M_r of Smad4/DPC4 (Fig. 4b). Note that the Smad4/DPC4 CTD elutes as a monomer from a gel-filtration column, which is consistent with residues N-terminal to the CTD contributing to homo-oligomerization¹⁵.

In principle, the full-length Smad4/DPC4 protein could assume oligomeric states other than a trimer but have a gel-filtration mobility approximating that of a trimer. However, our *in vivo* and *in vitro* results with the trimer-interface mutants, both with the CTD and the full-length proteins, indicates that the trimeric protein-protein interface in the crystals also participates in homo-oligomerization *in vivo*.

The Smad4/DPC4 CTD also supports hetero-oligomerization, as shown by the co-immunoprecipitation of overexpressed Smad4/DPC4 CTD and Smad2 CTD from COS cells¹⁵ (Fig. 3d), and by their association on electrophoresis in non-denaturing gels (data not shown). Furthermore, both the tumour-derived trimer-interface mutants and the L3-loop mutants abolished hetero-oligomerization between the CTDs of Smad4/DPC4 and Smad2 (Fig. 3d); results were similar with full-length Smad4/DPC4. As the L3-loop developmental mutation, which does not significantly affect homo-oligomerization, disrupts hetero-oligomer formation, the L3 loop may participate in hetero-oligomerization. Also, mutations preventing homo-oligomerization that also disrupt hetero-oligomerization indicate that the former is a prerequisite for the latter.

Although several models of hetero-oligomerization could explain our results, one that is suitable from a structural perspective is the formation of a heterohexamer between Smad4/DPC4 and Smad2 trimers. The trimer structure resembles a disk, with the L3 loops forming undulations on the face of the disk (Fig. 5a), so two disks could fit together face to face through their L3 loops (Fig. 5b), explaining why L3-loop mutations disrupt hetero-oligomerization. In this model, heterohexamer formation would

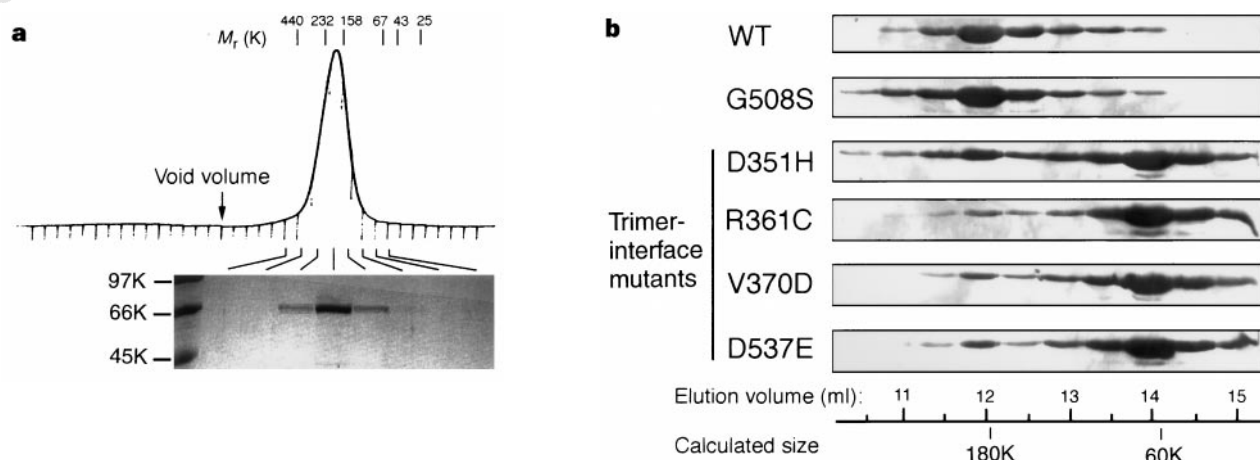


Figure 4 Size-exclusion chromatography showing that wild-type full-length Smad4/DPC4, but not tumour-derived mutants, has an apparent M_r consistent with that of a trimer. **a**, Recombinant Smad4/DPC4 protein, purified to near homogeneity, was applied to a Superdex-200 gel-filtration column, and it eluted at M_r 180K. Fractions were visualized with Coomassie blue staining. **b**, *In vitro*,

tumour-derived trimer-interface mutations disrupt homo-oligomerization, but a developmental mutation in the L3 loop does not. Gel-filtration fractions of partially purified wild-type and mutant Smad4/DPC4 proteins were analysed by immunoblotting with anti-Smad4/DPC4 antibody.

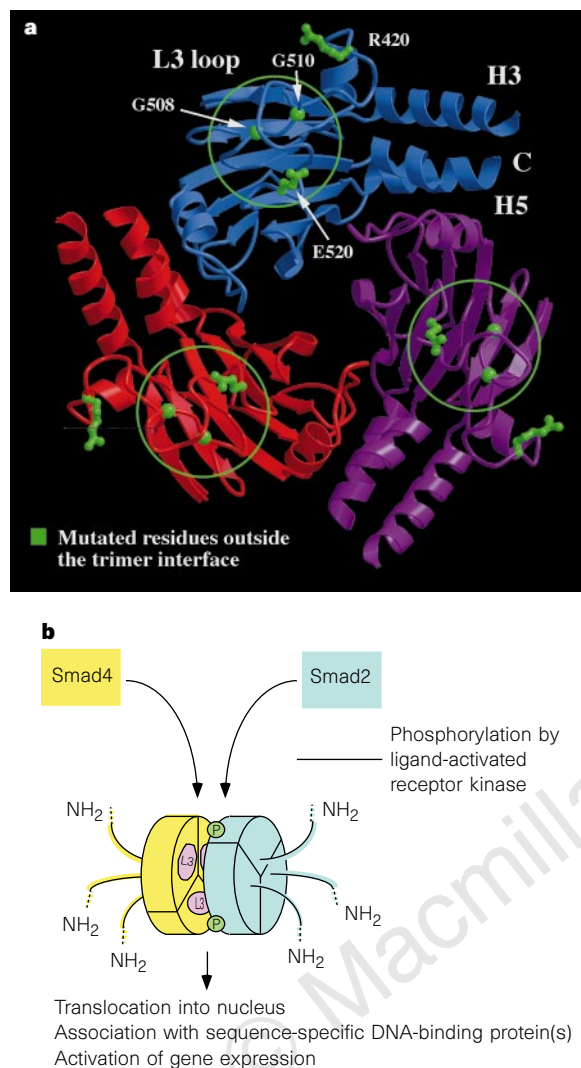


Figure 5 One face of the disk-like trimer structure may mediate hetero-oligomerization. **a**, Mutations outside the trimer interface map primarily to L3-loop residues, with the exception of Arg 420, which is outside the L3 loop. The face of the trimer shown is opposite to that shown in Fig. 3a. **b**, Model of hetero-oligomer formation depicting the Smad4/DPC4 and Smad2 CTD trimers as disks. The approximate positions of the Smad4/DPC4 L3 loops and of the Smad2 sites that get phosphorylated by the receptor kinase³⁰ are shown in yellow and green, respectively.

depend on homotrimer formation, explaining how tumorigenic mutations that disrupt homo-oligomerization can prevent the formation of a functional hetero-oligomeric complex and interfere with signal transduction. □

Methods

Protein expression and purification. Recombinant Smad4/DPC4 CTD, corresponding to residues 319–552, was overexpressed at room temperature in *Escherichia coli* using a pET vector (Novagen). The Smad4/DPC4 CTD in the soluble fraction of the *E. coli* lysate was partially purified on a Q-Sepharose column, concentrated by ultrafiltration, and further purified by gel-filtration chromatography (on a Superdex-75 column) and anion-exchange chromatography (Source-15Q column).

Crystallization. Crystals were initially grown at 4°C by the hanging-drop vapour-diffusion method by mixing the protein solution (10–15 mg ml⁻¹) with an equal volume of reservoir solution containing 100 mM MES, 25% monomethylether PEG5000 (mPEG5000) and 200 mM (NH₄)₂SO₄, pH 6.5.

Crystals suitable for diffraction were grown using streak seeding and macroseeding²⁰. The crystals form in the cubic space group *F*₄32, with *a* = *b* = *c* = 199.6 Å, and contain one molecule in the asymmetric unit.

Data collection and processing. Diffraction data were collected using an R-AXISIIIC imaging plate detector mounted on a Rigaku 200HB generator. Native 1 and derivative data were collected at 8°C, and native 2 data were collected at -170°C with a crystal flash-frozen in buffer containing 20% glycerol and 25% mPEG5000. Heavy-atom soaks were done in 50 mM HEPES, 25% mPEG, 160 mM (NH₄)₂SO₄, 100 mM NaCl, pH 6.1, containing one of the following heavy-atom solutions: 1.2 mM thimerosal (for 12 h), 3.0 mM (CH₃)₃PbOOCCH₃ (for 3 d), or 2.0 mM uranyl acetate (for 19 h).

MIR analysis, model building and refinement. The heavy-atom sites of the thimerosal derivative were determined by direct methods with the program SHELXS-90 (ref. 21), and the heavy-atom sites of the other derivatives were identified by difference Fourier methods. Initial MIR phases calculated with the program MLPHARE²² had a mean figure of merit of 0.62 to 3.2 Å, and were improved with solvent flattening and histogram matching with the program SQUASH²³. The MIR electron density maps had continuous electron density for most of the Smad4/DPC4 CTD polypeptide, apart from a 34-amino-acid region between helices H3 and H4. A model was built into MIR electron density maps with program O (ref. 24), refined by simulated annealing with the program X-PLOR²⁵, and checked by calculating X-PLOR omit maps in which 5–7% of the structure was deleted in each calculation; simulated annealing was used to reduce model bias. The refined model contains residues 319–543 of human Smad4/DPC4 and 129 water molecules. Residues 544–552 at the C terminus and residues 457–491 between helices H3 and H4 have no electron density in the maps—we presume that these regions are disordered in the crystals.

In vivo oligomerization assays. Full-length Smad4/DPC4 and Smad2, Smad4/DPC4 CTD encoding amino acids 294–552, and the Smad2 CTD encoding amino acids 248–467 were subcloned into the mammalian expression vector pCMV5. All Smad4/DPC4 point mutations were generated by a polymerase chain reaction (PCR)-based method and were confirmed by DNA sequencing. Mammalian COS-1 cells were transiently transfected with the indicated Flag- and HA-tagged constructs by the DEAE-dextran method. Two days after transfection, cells were incubated with 200 pM TGF-β1 for one hour. Cells were lysed, the lysate immunoprecipitated and immunoblotted as described⁹. Wash buffers contained 150 mM NaCl for all immunoprecipitations, except for the homo-oligomerization assay of full-length wild-type (WT) Smad4/DPC4 and point mutants, in which 250 mM NaCl gave better differentiation of WT and mutant activities.

In vitro oligomerization assays. Full-length Smad4/DPC4 proteins, both wild-type and point mutants, were overexpressed at room temperature in *E. coli* using a pET vector (Novagen). Smad4/DPC4 protein in the soluble fraction of the *E. coli* lysate was partially purified by ion-exchange chromatography on Q-Sepharose and by gel filtration on Superdex-200 in 50 mM Tris, 200 mM NaCl, 5 mM DTT, pH 8.0. Aliquots from fractions corresponding to *M*_r standards between 440K and 25K were taken for immunoblotting with a rabbit polyclonal antibody raised against the Smad4/DPC4 CTD. Results were visualized by western blot analysis and ECL detection (Amersham). WT full-length Smad4/DPC4 was also cloned as a glutathione *S*-transferase fusion protein and purified to near homogeneity over a glutathione column; the fusion protein was then cleaved with thrombin and the Smad4/DPC4 protein further purified by anion-exchange chromatography (Source-15Q column).

Received 22 January; accepted 8 May 1997.

- Hahn, S. A. et al. DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science* **271**, 350–353 (1996).
- Schutte, M. et al. DPC4 gene in various tumor types. *Cancer Res.* **56**, 2527–2530 (1996).
- Kim, S. K. et al. DPC4, a candidate tumor suppressor gene, is altered infrequently in head and neck squamous cell carcinoma. *Cancer Res.* **56**, 2519–2521 (1996).
- Nagatake, M. et al. Somatic *in vivo* alterations of the DPC4 gene at 18q21 in human lung cancers. *Cancer Res.* **56**, 2718–2720 (1996).
- Massague, J. TGF-β signalling: receptors, transducers, and mad proteins. *Cell* **85**, 947–950 (1996).
- Derynck, R. et al. Nomenclature: vertebrate mediators of TGFβ family signals. *Cell* **87**, 173 (1996).
- Derynck, R. & Zhang, Y. Intracellular signalling: the Mad way to do it. *Curr. Biol.* **6**, 1226–1229 (1996).
- Zhang, Y., Feng, X.-H., Wu, R.-Y. & Derynck, R. Receptor-associated Mad homologues synergize as effectors of the TGF-β response. *Nature* **383**, 168–172 (1996).
- Lagna, G., Hata, A., Hemmati-Brivanlou, A. & Massague, J. Partnership between SPC4 and SMAD proteins in TGF-β signalling pathways. *Nature* **383**, 832–836 (1996).
- Yingling, J. M. et al. Mammalian dwarfins are phosphorylated in response to transforming growth factor β and are implicated in control of cell growth. *Proc. Natl Acad. Sci. USA* **93**, 8940–8944 (1996).

11. Hoodless, P. A. *et al.* MADR1, a MAD-related protein that functions in BMP2 signaling pathways. *Cell* **85**, 489–500 (1996).
12. Eppert, K. *et al.* MADR2 maps to 18q21 and encodes a TGF β -regulated MAD-related protein that is functionally mutated in colorectal carcinoma. *Cell* **86**, 543–552 (1996).
13. Lechleider, R. J., de Caestecker, M. P., Dehejia, A., Polymeropoulos, M. H. & Roberts, A. B. Serine phosphorylation, chromosomal localization, and transforming growth factor- β signal transduction by human bsp-1. *J. Biol. Chem.* **271**, 17617–17620 (1996).
14. Liu, F. *et al.* A human Mad protein acting as a BMP-regulated transcriptional activator. *Nature* **381**, 620–623 (1996).
15. Hata, A., Lo, R. S., Wotton, D. & Massagué, J. Mutations increasing autoinhibition inactivate the tumour suppressors Smad2 and Smad4. *Nature* **388**, 82–87 (1997).
16. Baker, J. C. & Harland, R. M. A novel mesoderm inducer, Madr2, functions in the activin signal transduction pathway. *Genes Dev.* **10**, 1880–1889 (1996).
17. Thiagalingam, S. *et al.* Evaluation of candidate tumour suppressor genes on the chromosome 18 in colorectal cancers. *Nature Genet.* **13**, 343–346 (1996).
18. Sekelsky, J. J., Newfield, S. J., Raftery, L. A., Chartoff, E. H. & Gelbart, W. M. Genetic characterization and cloning of *Mothers against dpp*, a gene required for decapentaplegic function in *Drosophila melanogaster*. *Genetics* **139**, 1347–1358 (1995).
19. Savage, C. *et al.* *Caenorhabditis elegans* genes *smad-2*, *smad-3* and *smad-4* define a conserved family of transforming growth factor β pathway components. *Proc. Natl Acad. Sci. USA* **93**, 790–794 (1996).
20. Stura, E. A. & Wilson, I. A. In *Crystallization of Nucleic Acids and Proteins* (eds Ducruix, A. & Giegé, R.) 99–126 (Oxford Univ. Press, 1992).
21. Sheldrick, G. in *Patterson Interpretation and the Use of Macromolecular Delta-F Data* (Daresbury, UK, 1991).
22. Collaborative Computational Project, B. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
23. Zhang, K. Y. J. SQUASH-combining constraints for macromolecular phase refinement and extension. *Acta Crystallogr. D* **49**, 213–222 (1993).
24. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–110 (1991).
25. Brunger, A. T. *X-PLOR, a system for crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1991).
26. Kluwe, P. J. Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
27. Merritt, E. A. & Murphy, M. E. Raster3D Version 2.0: a program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).
28. Uchida, K. *et al.* Somatic *in vivo* alterations of the JV18-1 gene at 18q21 in human lung cancers. *Cancer Res.* **56**, 5583–5585 (1996).
29. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).
30. Macias-Silva, M. *et al.* MADR2 is a substrate of the TGF β receptor and its phosphorylation is required for nuclear accumulation and signalling. *Cell* **87**, 1215–1224 (1996).

Acknowledgements. This work was supported by the NIH, the Pew Charitable Trusts, the Arnold and Mabel Beckman Foundation, the Dewitt Wallace Foundation, the Samuel and May Rudin Foundation, and the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to N.P.P. (e-mail: nikola@xray2.mskcc.org). Coordinates are being deposited with the Brookhaven Protein Data Bank.

DNA-replication checkpoint control at the *Drosophila* midblastula transition

Ody C. M. Sibon, Victoria A. Stevenson & William E. Theurkauf

Department of Biochemistry and Cell Biology and the Institute for Cell and Developmental Biology, State University of New York at Stony Brook, Stony Brook, New York 11794-5215, USA

Embryogenesis is typically initiated by a series of rapid mitotic divisions that are under maternal genetic control¹. The switch to zygotic control of embryogenesis at the midblastula transition is accompanied by significant increases in cell-cycle length and gene transcription, and changes in embryo morphology^{2,3}. Here we show that mutations in the *grapes* (*grp*) checkpoint 1 kinase homologue⁴ in *Drosophila* block the morphological and biochemical changes that accompany the midblastula transition, lead to a continuation of the maternal cell-cycle programme, and disrupt DNA-replication checkpoint control of cell-cycle progression. The timing of the midblastula transition is controlled by the ratio of nuclei to cytoplasm (the nucleocytoplasmic ratio), suggesting that this developmental transition is triggered by titration of a maternal factor by the increasing mass of nuclear material that accumulates during the rapid embryonic mitoses^{5–9}. Our observations support a model for cell-cycle control at the midblastula transition in

which titration of a maternal component of the DNA-replication machinery slows DNA synthesis and induces a checkpoint-dependent delay in cell-cycle progression¹⁰. This delay may allow both completion of S phase and transcription of genes that initiate the switch to zygotic control of embryogenesis.

The rapid maternally controlled divisions that initiate embryogenesis in *Drosophila melanogaster* proceed without cytokinesis, and produce a syncytial blastoderm embryo¹¹. We used a screen for mutations that disrupt mitotic chromosome segregation¹² to identify a maternal-effect mutation, initially designated *fs(A)4*, that produced severe mitotic defects during the later syncytial divisions.

To analyse directly the effects of this mutation on spindle morphogenesis and cell-cycle progression, we micro-injected embryos derived from homozygous *fs(A)4* females with rhodamine-tubulin conjugates, and visualized microtubule reorganization and cell-cycle progression by time-lapse confocal microscopy^{13,14} (Fig. 1). During the final syncytial divisions before the midblastula transition (MBT) (divisions 11–13), the duration of interphase increases from 6 to 14 min (Table 1). Interphase 14 is prolonged, lasting over 60 min, during which time invaginating membranes surround the majority of nuclei to form a cellular blastoderm¹¹. Blastoderm cellularization is the first important morphogenetic event that requires zygotic transcription, and thus cytologically marks the *Drosophila* MBT⁵. Our *in vivo* analyses of the syncytial mitoses demonstrate that embryos derived from homozygous *fs(A)4* females do not show a significant increase in interphase length during syncytial divisions 11–13 (Fig. 1, Table 1). Furthermore, mutant embryos proceed through at least two extra syncytial cycles of spindle assembly and disassembly and nuclear envelope breakdown and formation after mitosis 13, and cellularization is never observed (Fig. 1; data not shown). The *fs(A)4* mutation thus blocks nearly all of the changes in cell-cycle dynamics and embryo morphology associated with the *Drosophila* MBT.

The changes in cell-cycle length at the *Drosophila* MBT are associated with changes in phosphorylation of the cyclin-dependent kinase Cdc2 (ref. 15). Termination of the rapid syncytial divisions normally occurs between 2 and 3 h after egg deposition, and is accompanied by accumulation of a low electrophoretic-mobility inhibitory tyrosine-phosphorylated form of Cdc2 (ref. 15) (form 4, Fig. 2a). In *fs(A)4* mutant embryos, this tyrosine-phosphorylated form of Cdc2 does not accumulate (Fig. 2a). Studies in yeast and vertebrates indicate that the Cdc2–cyclin complex is activated by the Cdc25 phosphatase, which removes the inhibitory tyrosine phosphates and thus drives the cell into mitosis¹⁶. In *Drosophila*, a dramatic decrease in the level of String, the *Drosophila* Cdc25 homologue, accompanies the increase in cell-cycle time at the MBT^{15,17}. In *fs(A)4* mutants, String protein levels do not decrease significantly during the later syncytial mitoses (Fig. 2b). The biochemical changes in Cdc2 that normally accompany the MBT are therefore blocked by the *fs(A)4* mutation.

These observations raise the possibility that the *fs(A)4* gene product slows the cell cycle by inducing a downregulation of the String phosphatase, thereby allowing accumulation of inhibitory tyrosine-phosphorylated forms of Cdc2. It is also possible that coordinate activation of the *Drosophila* wee 1 kinase, which is in part responsible for inhibitory tyrosine phosphorylation of Cdc2 (ref. 18), may also be involved at the MBT.

The MBT is characterized by a sudden increase in zygotic transcription. In *Drosophila*, the MBT is accompanied by transcription of segmentation genes that are expressed in spatially restricted domains. We therefore used whole-mount *in situ* hybridization to assay the onset of zygotic segmentation gene expression in *fs(A)4* mutant embryos. Before the MBT, the segmentation genes *runt*, *fushi tarazu* (Fig. 3) and *giant* (not shown) are expressed at relatively low levels over broad regions of the embryo¹⁹, and the early patterns of *runt*, *fushi tarazu* and *giant* expression are observed in *fs(A)4* mutants. After mitosis 13, these genes are normally expressed at

high levels in characteristic stripes¹⁹ (Fig. 3). In *fs(A)4* mutant embryos, by contrast, the post-MBT expression patterns of these genes are not observed (Fig. 3). These data indicate that qualitative changes in zygotic transcription at the MBT are disrupted by the *fs(A)4* mutation.

These data indicate that the *fs(A)4* mutation blocks nearly all of the changes in embryo morphology, cell-cycle dynamics and biochemistry, and transcriptional regulation that normally accompany the MBT. We therefore conclude that *fs(A)4* mutant embryos do not undergo the MBT, but continue to divide under maternal cell-cycle control.

To gain insight into the biochemical function of the gene identified by *fs(A)4*, we initiated detailed molecular and genetic analyses of this mutation. The *fs(A)4* mutation was generated in a P-element transposon mutagenesis screen²⁰. Reversion analyses were performed to determine whether transposon excision could restore *fs(A)4* gene function. Most of the P-element excision chromosomes recovered (19 of 26) were homozygous female fertile and complemented the original *fs(A)4* mutation²⁰. The other excision chromosomes were sterile in combination with the original *fs(A)4* allele and

when homozygous, suggesting that the transposon excision events had generated deletion alleles. We recovered sequences flanking the P element in the *fs(A)4* stock by the plasmid rescue technique, and used these sequences to localize the insertion to polytene region 36AB by *in situ* hybridization. Standard crosses to deficiency chromosomes were then used to confirm that the mutation was located in the 36AB region (data not shown). These results provided strong genetic evidence that the *fs(A)4* mutation was caused by a P-element transposon insertion into a gene in the 36AB region of chromosome 2.

Polytene region 36AB contains the previously identified maternal-effect lethal mutation *grapes* (*grp*)^{12,21}. We therefore tested the *fs(A)4* chromosome for the ability to complement *grp*; *fs(A)4* fails to complement a P-element-induced loss-of-function allele of *grp* (*grp*¹)^{12,21}. We then used genomic Southern blotting analyses to map the P-element insertion in *fs(A)4* to within 300 base pairs of the *grp*¹ P-element insertion (data not shown). Finally, northern blotting with a *grp* cDNA probe demonstrated that both of the two largest *grp* transcripts were undetectable in embryos derived from homozygous *fs(A)4* females⁴ (data not shown). Based on these

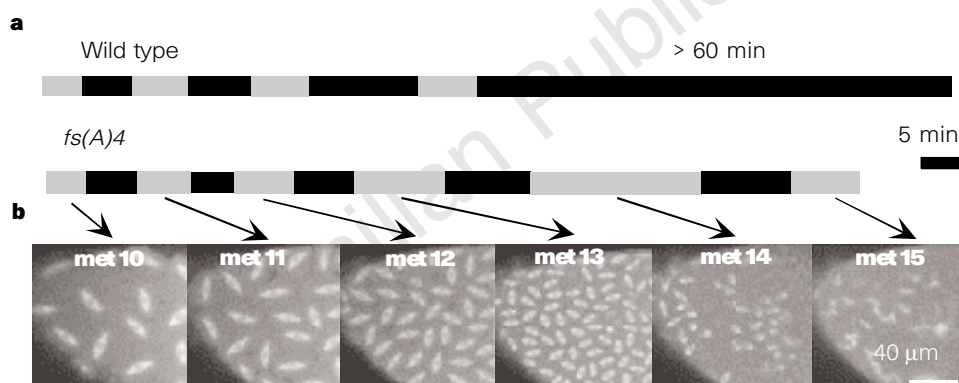


Figure 1 *fs(A)4* mutant embryos fail to undergo changes in cell-cycle timing and embryo morphology associated with the MBT. **a**, The bars show the average length of interphase (black) and mitosis (striped) in wild-type and *fs(A)4* mutant embryos (scale bar, 5 min; see Table 1). Wild-type embryos show a progressive increase in interphase length during syncytial divisions 11-13, followed by a prolonged interphase 14 of at least 60 min. In *fs(A)4* mutant embryos, interphase

length does not increase significantly during divisions 11-13; mitosis 13 is immediately followed by at least two additional syncytial cycles of spindle assembly/disassembly and nuclear envelope breakdown/assembly. **b**, Microtubule organization in a living *fs(A)4* mutant embryo at metaphase of syncytial cycles 10-15. A time-lapse animation of the continued syncytial divisions in *fs(A)4* mutations is available from <http://129.49.19.42/biochem/theurkauf.html>.

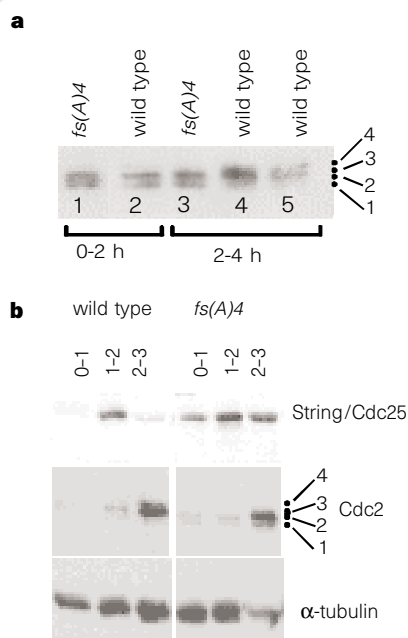


Figure 2 The *fs(A)4* mutation prevents inhibitory tyrosine phosphorylation of Cdc2 and String phosphatase downregulation at the MBT. **a**, Mutant and wild-type embryos 0-2 h old contain Cdc2 isoforms 1-3 (see ref. 15). Wild-type embryos 2-4 h old accumulate a low-mobility tyrosine-phosphorylated form of Cdc2 (form 4, lanes 4,5). Mutant embryos 2-4 h old accumulate only trace amounts of this form of Cdc2 (lane 3). **b**, Wild-type embryos show an increase in String levels during the first 2 h of development (String/Cdc25, 0-1 and 1-2 h), and a dramatic decrease in String levels as the cell cycle slows at the MBT (2-3 h time point)¹⁵. The *fs(A)4* embryos show the normal increases in String levels between 0 and 2 h, but do not dramatically downregulate String between 2 and 3 h. The middle panels show a duplicate blot probed for Cdc2. The presence of tyrosine-phosphorylated Cdc2 form 4 correlates precisely with the decline in String levels in wild-type embryos. The bottom panels show a duplicate western blot probed for α -tubulin.

molecular and genetic studies, we conclude that *fs(A)4* is a null or very strong loss-of-function allele of *grp*, which we designated *grp^{fs(A)4}*.

Loss of *grp* function has been reported to cause metaphase arrest at division 13 (ref. 21). We observed that *grp^{fs(A)4}* mutant embryos continue through rapid syncytial mitoses and fail to cellularize. To determine whether the continued syncytial divisions and failure to undergo the MBT were specific to the *grp^{fs(A)4}* allele, or were caused by a background mutation on the *grp^{fs(A)4}* chromosome, we analysed embryos derived from *grp^{fs(A)4}/grp¹*, *grp^{fs(A)4}/Df(2L)H10*, and *grp¹/grp¹* females. The progeny from females of each of these three genotypes showed defects in cell-cycle timing and cellularization that were indistinguishable from those observed in embryos derived from homozygous *grp^{fs(A)4}* females (data not shown). To determine whether these defects were induced by injection of the specific fluorescent probes used in our analyses, we examined embryos injected with rhodamine-tubulin, rhodamine-actin, the chromatin marker oli-green, and a mixture of oli-green and rhodamine tubulin (see Methods). We also assayed cell-cycle progression in uninjected embryos by time-lapse differential interference contrast microscopy. In all cases, we observed identical truncated syncytial cell cycles, additional syncytial mitoses, and a failure of blastoderm cellularization (data not shown). We therefore conclude that the phenotypes described here are not specific to the *grp^{fs(A)4}* allele.

In addition to the cell-cycle defects we report, embryos derived from *grp* mutant females show mitotic spindle and midbody defects during the later syncytial divisions²¹. We observe cytologically identical defects in spindle assembly in wild-type embryos treated with the DNA-synthesis inhibitor aphidicolin or exposed to X-rays, suggesting that these defects are a secondary consequence of the *grp* mutation. We believe that these mitotic defects reflect activation of a *grp*-independent pathway that disrupts the centrosome and blocks spindle function when mitosis is initiated in the presence of

damaged or incompletely replicated DNA (O.C.M.S. and W.E.T., manuscript in preparation). Activation of this pathway leads to spindle defects that seem to trigger the spindle checkpoint, leading to increases in the duration of mitosis during the later syncytial divisions in *grp^{fs(A)4}* embryos (Fig. 1). These observations and our analyses lead us to conclude that the primary function of the *grp* gene product is to control cell-cycle progression in the early embryo, and that loss of *grp* function leads to continued rapid syncytial mitoses.

We believe that several factors may have obscured the extra syncytial mitoses in *grp* mutants during earlier analyses²¹. Mitosis 13 in *grp* mutant embryos is prolonged, and the delay in the onset of anaphase could be mistaken for metaphase arrest (Fig. 1, Table 1). In addition, mitosis 14 and 15 are longer than the intervening interphases, leading to a very high mitotic index in older mutant embryos. Finally, chromosome segregation fails during cycles 13–15, so nuclear density does not increase. Analyses of fixed material and limited *in vivo* imaging could therefore lead to the conclusion that mutant embryos arrest in metaphase of division 13 (ref. 21), whereas the embryos actually proceed through additional syncytial mitoses.

The *grp* locus encodes a putative serine–threonine kinase that is most closely related to the check-1 kinase gene (*chk1*)⁴ in *Schizosaccharomyces pombe*. In fission yeast, the Chk1 protein is in a checkpoint pathway that inhibits cell-cycle progression in response to DNA damage or defects in DNA replication (refs 22, 23 and references therein). To determine whether Grp also lies in a replication checkpoint pathway, we analysed cell-cycle progression directly in living embryos after treatment with the DNA-synthesis inhibitor aphidicolin. For these studies, embryos were co-injected with rhodamine-tubulin and levels of aphidicolin that inhibit DNA synthesis by more than 95% (ref. 24). Embryos were then monitored by time-lapse confocal microscopy, and the time until

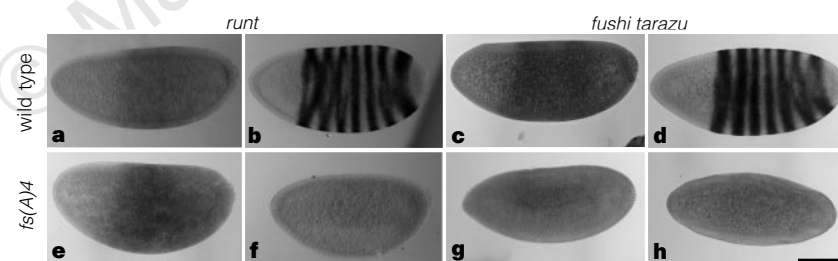


Figure 3 Expression pattern of early zygotic transcripts is impaired in *fs(A)4* mutants. Wild-type and *fs(A)4* mutant embryos were assayed for expression of early zygotic transcripts from the *runt* and *fushi tarazu* genes by whole-mount *in situ* hybridization. Wild-type embryos before cell cycle 13 show a broad expression pattern of both *runt* (a) and *fushi tarazu* (c). After metaphase 13, the

7-stripped expression pattern of *runt* (b) and *fushi tarazu* (d) are clearly visible¹⁹. The *fs(A)4* mutant embryos show the early broad expression patterns of both *runt* (e) and *fushi tarazu* (g), although the latter is less pronounced than in wild-type embryos (compare g and c). By contrast, the 7-stripped patterns of expression characteristic of the MBT were never observed in *fs(A)4* embryos (f, h).

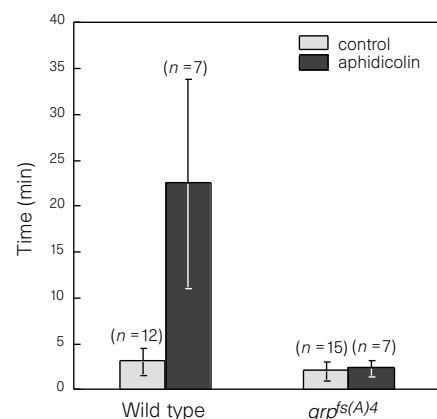


Figure 4 The *grp* gene is required to delay mitosis in response to the DNA-synthesis inhibitor aphidicolin. Wild-type embryos show a significant delay in initiation of mitosis in response in aphidicolin; *grp^{fs(A)4}* mutant embryos do not delay nuclear-envelope breakdown but proceed directly into mitosis despite an apparent block in completion of S phase. Embryos were injected with rhodamine-tubulin or a combination of rhodamine-tubulin and aphidicolin, and the time before nuclear-envelope breakdown was measured. The bar graphs represent data from divisions 11 and 12, and the s.d. is indicated. The number of live recordings scored is given in brackets.

Table 1 Cell-cycle timing in wild-type and *fs(A)/4* mutant embryos

Cell cycle stage	Wild-type average time (min)	Number	<i>fs(A)/4</i> average time (min)	Number
M 10	5.5 (0.7)	8	5.6 (0.4)	6
I 11	6.1 (1.1)	9	4.7 (1.0)	8
M 11	5.3 (0.8)	10	5.4 (0.7)	10
I 12	8.7 (0.9)	10	6.1 (1.1)	11
M 12	5.9 (1.0)	10	6.1 (1.1)	12
I 13	14.0 (1.3)	9	7.3 (1.6)	12
M 13	6.2 (1.0)	9	10.4 (2.3)	10
I 14	70.3 (3.8)	3	11.8 (3.9)	8
M 14	—	—	19.4 (5.9)	3

Mean (s.d.) duration of interphase (I) and mitosis (M) stages.

nuclear-envelope breakdown was measured (see Methods). In wild-type embryos during divisions 11 and 12, aphidicolin causes a mean fivefold delay in nuclear-envelope breakdown (Fig. 4). By contrast, *grp* mutant embryos do not delay nuclear-envelope breakdown in response to aphidicolin treatment, but proceed immediately into mitosis (Fig. 4). These observations indicate that a replication checkpoint pathway functions in the early embryo, and that the *grp* gene is essential to this pathway. The *grp* gene is expressed throughout the life cycle of the fly⁴. We have found that the postembryonic development of flies homozygous for the *grp* mutation is sensitive to hydroxyurea²⁵, a feature characteristic of mutations affecting replication checkpoint function (data not shown). The Grp replication checkpoint pathway may therefore serve a non-essential function through most of the life cycle of the fly, but is essential at the MBT.

Our analyses support a simple DNA-replication checkpoint model for cell-cycle control at the MBT¹⁰. In this model, a free-running cell-cycle oscillator drives the rapid divisions that initiate embryogenesis. During the first 10 divisions, maternally supplied components of the DNA synthesis/initiation machinery are in excess, and S phase is completed before the oscillator triggers mitosis. During divisions 10–13, in contrast, a component of the maternal replication machinery is titrated by the increasing mass of nuclear DNA and becomes rate limiting. DNA synthesis therefore slows, and S phase cannot be completed before the free-running cell-cycle oscillator would normally trigger mitosis. A Grp kinase-dependent checkpoint pathway is therefore activated and delays M phase until DNA synthesis is complete. This model is consistent with regulation of the MBT by the nucleocytoplasmic ratio, explains the absence of a G2 phase and the increase in S phase during the later syncytial mitoses, and accounts for the cell-cycle timing defects observed in *grp* mutant embryos. Cell-cycle progression in extracts of *Xenopus* embryos is also dependent on DNA replication¹⁰, raising the possibility that replication checkpoint control of the MBT is evolutionarily conserved.

The Grp protein could act directly in pathways that induce high-level zygotic transcription, blastoderm cellularization, and termination of syncytial division at the MBT. However, we favour a model in which the Grp protein functions specifically to control cell-cycle progression in anticipation of the MBT, and the failure to undergo later aspects of MBT in mutant embryos is a consequence of earlier defects in cell-cycle regulation. Consistent with this hypothesis, production of full-length primary transcripts from genes with large transcription units requires the wild-type increases in interphase length that precede the MBT^{26,27}. Further, embryos treated with the transcriptional inhibitor α -amanitin do not cellularize after mitosis 13, but proceed through an additional, 14th syncytial mitosis¹⁷. Finally, the high-level spatially patterned expression of several early zygotic genes is severely impaired or blocked in *grp* mutant embryos (Fig. 3). These observations suggest a model in which checkpoint-dependent inhibition of cell-cycle progression during divisions 10–13 allows complete transcription of zygotic genes, which in turn

act directly to terminate the maternal cell-cycle programme and trigger the transition to zygotic control of embryogenesis. □

Methods

In vivo cytology. For *in vivo* analyses of the early embryonic divisions, females of the specified genotypes were mated to wild-type Oregon R males, and the resulting embryos were injected with rhodamine-tubulin (Molecular Probes), rhodamine-actin, or a combination of rhodamine-tubulin and the chromatin marker oli-green (Molecular Probes). Mitosis and cell-cycle progression in injected embryos were assayed by time-lapse confocal microscopy as described for *Drosophila* oocytes²⁸. To quantify cell-cycle phases, interphase was defined by the presence of a nuclear envelope that excluded injected fluorescent protein conjugates, and mitosis was defined by the absence of a nuclear envelope and the presence of spindle structures. To determine the effects of inhibiting DNA synthesis on cell-cycle progression, embryos were injected with a mixture of rhodamine-tubulin (5 mg ml⁻¹) and aphidicolin (100 μ g ml⁻¹). Microinjection of aphidicolin at this concentration leads to more than 95% inhibition of DNA synthesis²⁴. Breakdown of the nuclear envelope was indicated by entry of rhodamine-tubulin and spindle assembly. Microinjection could not be controlled precisely with respect to cell-cycle phase, so these measurements were standardized as follows. On completion of mitosis, centrosomes duplicate and move to opposite poles of the nuclei, where they remain until nuclear-envelope breakdown and mitosis. We have found that the time between completion of centrosome migration and nuclear-envelope breakdown is consistent from embryo to embryo for a given cell cycle (Fig. 3). Completion of centrosome migration is therefore a consistent cytological marker for mid-interphase. We therefore injected embryos before the centrosomes had completed their movement to the poles, and determined the time between completion of migration and nuclear-envelope breakdown.

In situ hybridization. *In situ* hybridization for *runt*, *fushi tarazu* and *giant* transcripts was performed as described²⁹. Identical probes, hybridization conditions and colour development times were used for wild-type and mutant embryos.

Western blots. For western blot analyses, total protein from staged embryos was subjected to SDS polyacrylamide gel electrophoresis, transferred to membranes, and probed for Cdc2 or String as described¹⁵. As a loading control, duplicate blots were probed with a monoclonal anti- α -tubulin antibody.

Received 24 March; accepted 7 May 1997.

- Gilbert, S. F. *Developmental Biology* 3rd edn, 75–114 (Sinaur, Sunderland, MA, 1991).
- Newport, J. & Kirschner, M. A major developmental transition in early *Xenopus* embryos: I. characterization and timing of cellular changes at the midblastula stage. *Cell* **30**, 675–686 (1982).
- Foe, V. E., Odell, G. M. & Edgar, B. A. In *The Development of Drosophila melanogaster* (ed Bate, M. & Martinez Arias, A.) 149–300 (Cold Spring Harbor Laboratory Press, New York, 1993).
- Fogarty, P. et al. *grp*, a *Drosophila* gene with homology to the *S. pombe* *chk1/rad27* checkpoint gene, is required for the fidelity of the late syncytial divisions. *Curr. Biol.* (in the press).
- Edgar, B. A., Kiehle, C. P. & Schubinger, G. Cell cycle control by the nucleocytoplasmic ratio in early *Drosophila* development. *Cell* **44**, 365–372 (1986).
- Yasuda, G. K., Baker, J. & Schubinger, G. Temporal regulation of gene expression in the blastoderm *Drosophila* embryo. *Gen. Dev.* **5**, 1800–1812 (1991).
- Pritchard, D. K. & Schubinger, G. Activation of transcription in *Drosophila* embryos is a gradual process mediated by the nucleocytoplasmic ratio. *Gen. Dev.* **10**, 1131–1142 (1996).
- Clute, P. & Masui, Y. Regulation of the appearance of division asynchrony and microtubule-dependent chromosome cycles in *Xenopus laevis* embryos. *Dev. Biol.* **171**, 273–285 (1995).
- Kane, D. A. & Kimmel, C. B. The zebrafish midblastula transition. *Development* **119**, 447–456 (1993).
- Dasso, M. & Newport, J. W. Completion of DNA replication is monitored by a feedback system that controls the initiation of mitosis *in vitro*: studies in *Xenopus*. *Cell* **61**, 811–823 (1990).
- Foe, V. E. & Alberts, B. M. Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in *Drosophila* embryogenesis. *J. Cell Sci.* **61**, 31–71 (1983).
- Sullivan, W., Fogarty, P. & Theurkauf, W. E. Mutations affecting the cytoskeletal organization of syncytial *Drosophila* embryos. *Development* **118**, 1245–1254 (1993).
- Matthies, H. J., McDonald, H. B., Goldstein, L. S. & Theurkauf, W. E. Anastral meiotic spindle morphogenesis: role of the non-claret disjunctional kinesin-like protein. *J. Cell Biol.* **341**, 455–464 (1996).
- Kellogg, D. R., Mitchison, T. J. & Alberts, B. M. Behavior of microtubules and actin filaments in living *Drosophila* embryos. *Development* **103**, 675–686 (1988).
- Edgar, B. A., Sprenger, F., Duronio, R. J., Leopold, P. & O'Farrell, P. H. Distinct molecular mechanisms regulate cell cycle timing at successive stages of *Drosophila* embryogenesis. *Gen. Dev.* **8**, 440–452 (1994).
- Elledge, S. J. Cell cycle checkpoint: preventing an identity crisis. *Science* **274**, 1664–1672 (1996).
- Edgar, B. A. & Datar, S. A. Zygotic degradation of two maternal Cdc25 mRNAs terminates *Drosophila*'s early cell cycle program. *Gen. Dev.* **10**, 1966–1977 (1996).
- Campbell, S. D., Sprenger, F., Edgar, B. A. & O'Farrell, P. H. *Drosophila* Wee 1 kinase rescues fission yeast from mitotic catastrophe and phosphorylates *Drosophila* Cdc2 *in vitro*. *Mol. Biol. Cell* **6**, 1333–1347 (1995).
- Klinger, M. & Gergen, J. P. Regulation of *runt* transcription by *Drosophila* segmentation genes. *Mech. Dev.* **43**, 3–19 (1993).
- Cooley, L., Kelley, R. & Spradling, A. Insertional mutagenesis of the *Drosophila* genome with single P elements. *Science* **239**, 1121–1128 (1988).

21. Fogarty, P., Kalpin, R. F. & Sullivan, W. The *Drosophila* maternal-effect mutation *grapes* causes a metaphase arrest at nuclear cycle 13. *Development* **120**, 2131–2142 (1994).
22. Francesconi, S., Grenon, M., Bouvier, D. & Baldacci, G. p56^{lck} protein kinase is required for the DNA replication checkpoint at 37 °C in fission yeast. *EMBO J.* **16**, 1332–1341 (1997).
23. Walworth, N. C. & Bernards, R. rad-dependent response of the chk1-encoded protein kinase at the DNA damage checkpoint. *Science* **271**, 353–356 (1996).
24. Raff, J. W. & Glover, D. M. Nuclear and cytoplasmic cycles continue in *Drosophila* embryos in which DNA synthesis is inhibited by aphidicolin. *J. Cell Biol.* **107**, 2009–2019 (1988).
25. Banga, S. S., Shenkar, R. & Boyd, J. B. Hypersensitivity of *Drosophila* *mei-41* mutants to hydroxyurea is associated with reduced mitotic chromosome stability. *Mutant. Res.* **163**, 157–165 (1986).
26. Shermoen, A. W. & O'Farrell, P. H. Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. *Cell* **67**, 303–310 (1991).
27. Ruden, D. M. & Jackle, H. Mitotic delay dependent survival identifies components of cell cycle control in the *Drosophila* blastoderm. *Development* **121**, 63–73 (1995).
28. Theurkauf, W. E. in *Methods of Cell Biology: Drosophila melanogaster: Practical Uses in Cell and Molecular Biology* (eds Goldstein, L. S. B. & Fryberg, E.) 489–505 (Academic, New York, 1994).
29. Tautz, D., Pfeifle, C. A non-radioactive in situ hybridization method for the localization of specific RNAs in *Drosophila* embryo's reveals translational control of the segmentation gene *hunchback*. *Chromosoma* **98**, 81–85 (1989).

Acknowledgements. We thank W. Sullivan, S. Campbell, B. Holdener, J. Kramer, N. Hollingsworth and B. Eggen for comments on the manuscript; B. Edgar and P. O'Farrell for discussions; P. Gergen for probes and advice on whole-mount *in situ* hybridization; P. Fogarty and W. Sullivan for sharing unpublished data and providing *grp* cDNA clones; R. S. Hawley and colleagues at the University of California at Davis for providing their collection of P-element-associated maternal-effect lethal mutations; and B. Edgar for anti-Cdc2 antibodies and advice on cell-cycle control during early embryogenesis. This work was supported by a fellowship from the NWO to O.C.M.S. and grants from the NIH and American Cancer Society to W.E.T.

Correspondence and requests for materials should be addressed to W.E.T. (e-mail: theurkauf@mcbgsg.bio.sunysb.edu).

Structure of the multimodular endonuclease *FokI* bound to DNA

David A. Wah, Joel A. Hirsch*, Lydia F. Dörner†, Ira Schildkraut† & Aneel K. Aggarwal*

Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, USA

† New England Biolabs, 32 Tozer Road, Beverly, Massachusetts 01915, USA

FokI is a member of an unusual class of bipartite restriction enzymes that recognize a specific DNA sequence and cleave DNA nonspecifically a short distance away from that sequence^{1–3}. Because of its unusual bipartite nature, *FokI* has been used to create artificial enzymes with new specificities^{4–7}. We have determined the crystal structure at 2.8 Å resolution of the complete *FokI* enzyme bound to DNA. As anticipated, the enzyme contains amino- and carboxy-terminal domains corresponding to the DNA-recognition and cleavage functions, respectively. The recognition domain is made of three smaller subdomains (D1, D2 and D3) which are evolutionarily related to the helix–turn–helix-containing DNA-binding domain of the catabolite gene activator protein CAP⁸. The CAP core has been extensively embellished in the first two subdomains, whereas in the third subdomain it has been co-opted for protein–protein interactions. Surprisingly, the cleavage domain contains only a single catalytic centre, raising the question of how monomeric *FokI* manages to cleave both DNA strands. Unexpectedly, the cleavage domain is sequestered in a 'piggyback' fashion by the recognition domain. The structure suggests a new mechanism for nuclease activation and provides a framework for the design of chimaeric enzymes with altered specificities.

FokI exists as a monomer and recognizes an asymmetric DNA sequence 5'-GGATG-3', cleaving DNA phosphodiester groups 9 and 13 base pairs (bp) away from the recognition site (Fig. 1). The structure shows the enzyme approaching DNA from the major-groove side, where it appears to surround the DNA (Fig. 2). Subdomain D1 of the recognition domain covers the DNA major

groove, recognizing base pairs at the 3' end of the recognition sequence (GGATG). D1 is a compact structure of eight helices ($\alpha 1$ to $\alpha 8$), two loops (L1 and L2), a β -sheet ($\beta 1$ to $\beta 3$) and an N-terminal arm (Fig. 3a). Helices $\alpha 4$, $\alpha 5$ and $\alpha 6$ and loops L1 and L2 make up the modified helix–turn–helix (HTH) motif. The short helices $\alpha 4$ and $\alpha 5$ share the same helical axis, as if they were part of a single α -helix into which L1 has been inserted. Together, $\alpha 4$ and $\alpha 5$ ($\alpha 4/5$) form the first helix of the HTH motif; $\alpha 6$ is the second helix (also known as the recognition helix) which lies against the DNA major groove. The turn expected between $\alpha 4/5$ and $\alpha 6$ is replaced by L2. Both loops reach out like fingers from the helices to make base-specific contacts with the DNA. D1 connects to D2, which contacts base pairs at the 5' end of the recognition sequence (GGATG). D2 is an extended, triangular-shaped structure of six helices ($\alpha 1$ to $\alpha 6$), a β -sheet ($\beta 1$, $\beta 2$, $\beta 5$), a β -hairpin ($\beta 3$, $\beta 4$), and a short loop L1 (Fig. 3a). As in D1, the basic HTH motif is extensively modified. $\alpha 2$ and $\alpha 5$ comprise the two α -helices of the HTH motif, with $\alpha 5$ lying in the DNA major groove. The 'turn' is replaced by loop L1, a pair of antiparallel helices $\alpha 3$ and $\alpha 4$, and a short segment T1 connecting $\alpha 4$ to $\alpha 5$. D3 is most similar to CAP and related proteins such as histone H5, HNF-3 γ and the biotin operon repressor (BirA)^{8–11} (Fig. 3a). Despite this similarity, D3 barely touches the DNA. Remarkably, its recognition helix lies outside the DNA major groove and is used primarily to piggyback the cleavage domain onto the recognition domain. Thus, we observe an almost identical HTH motif being used in a different context. In the CAP-related protein–DNA complexes, the motif is used to recognize the DNA bases, whereas in D3 the motif is used primarily to mediate protein–protein interactions.

As already noted, subdomains D1 and D2 make almost all of the base-specific contacts with the DNA. The DNA maintains a B-form DNA conformation without any major bends or kinks. Interactions between D1 and DNA occur along four segments of the subdomain. The N-terminal arm, loop L1, and the recognition helix $\alpha 6$ form specific contacts in the major groove, whereas loop L2 makes contacts in the minor groove. D2, on the other hand, contacts DNA exclusively in the major groove by $\alpha 4$ and T1 of the 'turn', and the recognition helix $\alpha 5$. The recognition helices of D1 and D2 approach DNA differently. The D1 recognition helix is most similar to the canonical form, packing against the major groove with its α -helical axis roughly perpendicular to the DNA axis¹². In contrast, the recognition helix of D2 juts away from the DNA, its α -helical axis tilted by $\sim 35^\circ$ with respect to the plane of the base pairs. The 'angle of attack' is reminiscent of the way the α -helices of the zinc-fingers of protein Zif268 approach DNA¹³. As a unit, the HTH of D2 is flipped 180° with respect to the HTH of D1, analogous to the inverse orientations of the HTH motifs of the homeodomain and the POU-specific domain of the Oct-1 POU domain¹⁴.

The protein–DNA interactions are extensive. For instance, almost all of the hydrogen-bond donor and acceptor groups in the major groove of the recognition sequence are involved in direct contacts with the protein. This complementarity at the protein–DNA interface ensures that only the *FokI* recognition sequence can form all of the interactions. The three guanines of the recognition sequence (GGATG) make bidentate hydrogen bonds with arginine and lysine residues, and the adenines of the third and fourth base

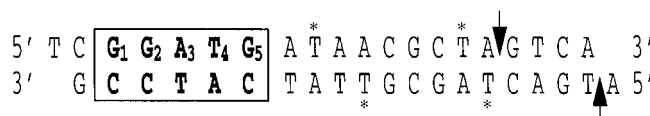


Figure 1 Sequence of the 20-bp DNA fragment used to co-crystallize *FokI*. The recognition sequence is numbered and the sites of cleavage are indicated by arrows. Asterisks denote the thymine residues substituted by iodouracils in the derivative Iodo (Table 1). This oligomer can be cleaved by the enzyme (data not shown).

* Present addresses: Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511, USA (J.A.H.); Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York, New York 10029, USA (A.K.A.).

pairs (GGATG) form bidentate hydrogen bonds with glutamine and asparagine residues (see Fig. 3b,c for details). We also observe novel interactions involving an aromatic tryptophan residue (Trp 105), which is wedged against the non-polar edges of three consecutive pyrimidines of base pairs 4 to 6 (GGATGA) in the major groove with impressive van der Waals complementarity (Fig. 3c). To our knowledge, this is the first time that a tryptophan residue has been observed in such a context, and it appears to represent a particularly efficient way of specifying consecutive pyrimidine residues in the major groove.

The cleavage domain has a structure remarkably similar to a subunit of the dimeric endonuclease *Bam*HI (ref. 15) (Fig. 4). The resemblance was unexpected as the two enzymes lack sequence similarity. Furthermore, *Bam*HI exists as a dimer and cleaves at a specific DNA site, and *Fok*I exists as a monomer and cleaves at a nonspecific site¹⁵. The *Fok*I cleavage domain has a compact α/β architecture consisting of a mixed six-stranded β -sheet (β 1 to β 6) with α -helices (α 1, α 2, α 3, α 6 and α 4, α 5) on each side (Fig. 4). The

similarity with *Bam*HI allows us to align the active sites of the two enzymes, which occur at the end of a β -meander formed by strands β 1, β 2, β 3 in *Fok*I and strands β 3, β 4, β 5 in *Bam*HI (Fig. 4). Remarkably, the three residues Asp 450, Asp 467 and Lys 469 that are important for catalysis in *Fok*I (ref. 16) overlap exactly with the catalytic residues Asp 94, Glu 111 and Glu 113 of *Bam*HI (ref. 15). The identification of only a single catalytic centre raises the question of how monomeric *Fok*I manages to cleave both DNA strands. Consistent with the notion of a single active site, it has been shown that changing either Asp 450 or Asp 467 to an alanine in *Fok*I results in the loss of cleavage of both DNA strands¹⁶. Thus, in order for the enzyme to cleave both strands, we suggest that it either dimerizes on DNA or that its cleavage domain flips between two orientations, cleaving first one DNA strand and then the other.

Probably the most surprising feature of our complex is the position of the cleavage domain (Fig. 2). Our expectation was that the cleavage domain would be located in a major groove, cleaving DNA one turn away from the recognition sequence.

Table 1 Crystallographic analysis

	Native	Hg 1*	Hg 2*	Hg 3*	Au	Sm	Iodo
Detector	CCD	CCD	IPs	CCD	CCD	CCD	IPs
Resolution (Å)	2.8	3.0	3.0	3.0	3.0	4.1	3.0
Unique reflections	26,243	17,014	18,759	20,950	16,268	7,046	20,397
Data coverage (%)	98.5	78	87	97	75	82	94
R_{merge} (%)†	7.5	6.2	8.0	7.5	4.6	2.8	8.2
Phasing statistics:							
R_{iso} (%)‡		9.0	8.1	10.2	7.1	7.7	14.4
Phasing power (isomorphous)§		1.74	1.73	1.41	1.35	1.45	
Phasing power (anomalous)¶			1.38	1.35	1.85	1.68	
Refinement statistics:							
(8 to 2.8 Å)							
Reflections, $F > 2\sigma(F)$	24,931						
R -factor (%)	21.4						
R -free (%)‡	29.6						
R.m.s. bond lengths (Å)	0.011						
R.m.s. bond angles (°)	1.7						

CCD, charge-coupled device; IPs, imaging plates.

* The mercurial derivatives differ in the times of soak and the relative occupancies and positions of the heavy-atom sites.

† $R_{\text{merge}} = \sum |I_{\text{obs}} - \langle I \rangle| / \sum \langle I \rangle$, calculated for all data.

‡ $R_{\text{iso}} = \sum |F_{\text{p}} - F_{\text{ph}}| / \sum F_{\text{p}}$, where F_{p} and F_{ph} are the structure factor amplitudes of the native and derivative data, respectively.

§ Phasing power (isomorphous) = $\langle |F_{\text{h}}| \rangle / \text{r.m.s.}(\epsilon)$, where $\langle |F_{\text{h}}| \rangle$ is the mean calculated amplitude for the heavy-atom model and ϵ is the lack of closure error.

¶ Phasing power (anomalous) = $\langle |\Delta F_{\text{h}}| \rangle / \text{r.m.s.}(\epsilon')$, where $\langle |\Delta F_{\text{h}}| \rangle$ is the mean calculated Bijvoet difference from the heavy-atom model and ϵ' is the lack of closure for anomalous differences.

‡ Calculated with 3% of the reflection data excluded from refinement.

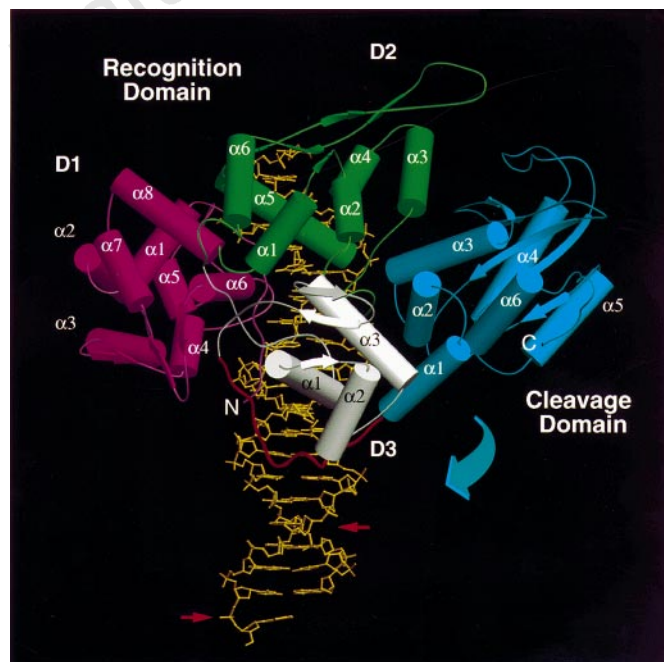


Figure 2 The complete *Fok*I enzyme (579 amino acids) bound to DNA. The recognition domain consists of three smaller subdomains D1, D2 and D3 shown in purple, green and white, respectively. The cleavage domain is shown in blue and the linker segment connecting it to the recognition domain is shown in red. The DNA (yellow) is the 20-bp DNA fragment shown in Fig. 1. Small arrows indicate the sites of cleavage. The large arrow conveys the proposed motion of the cleavage domain toward the cleavage site. The structure assignment is as follows: D1, α 1(17-25), α 2(32-45), α 3(49-58), β 1(66-67), α 4(68-72), L1(73-86), α 5(87-91), L2(92-103), α 6(104-116), β 2(120-123), β 3(128-131), α 7(133-140), α 8(146-156); D2, α 1(160-168), β 1(175-176), α 2(177-182), L1(183-196), α 3(197-205), α 4(211-218), T1(219-221), α 5(222-236), β 2(240-242), β 3(245-248), β 4(258-261), β 5(264-267), α 6(269-277); D3, α 1(303-319), β 1(323-324), α 2(325-335), α 3(341-353), β 2(358-361), β 3(364-367); linker (373-387); cleavage domain, α 1(389-400), α 2(406-414), α 3(421-434), β 1(439-442), β 2(451-454), β 3(462-469), α 4(479-491), β 4(514-521), α 5(528-539), β 5(543-547), α 6(548-560), β 6(575-576).

Instead, we find that the cleavage domain is packed alongside the recognition domain and is prevented from associating with the DNA-cleavage site. This is in agreement with footprinting studies on *FokI* with DNA which show a lack of protection at the cleavage site^{16–18}. From our structure, the cleavage domain can be brought to the major groove for cleavage by simple rotations around the linker segment. The sequestering of the cleavage domain may explain how *FokI* manages to regulate its cleavage activity until it is required. The compact configuration we observe may be the state in which the enzyme scans the DNA looking for the correct site, and upon finding it, in the presence of Mg^{2+} , the cleavage domain swings over to the major groove for DNA cleavage. The recognition domain sequesters the cleavage domain through an extensive set of protein–protein interactions (Fig. 5). These interactions are primarily electrostatic and mediated by subdomains D2 and D3 of the recognition domain, the linker segment, and helices $\alpha 1$ and $\alpha 3$ and strand $\beta 1$ of the cleavage domain.

For the enzyme to become activated for cleavage, these protein–protein interactions would need to be replaced by intermolecular interactions with the DNA. The binding of Mg^{2+} may provide part of the free energy for the configurational switch of the enzyme by promoting the formation of the active site. Evidence for sequestration of the cleavage domain by the recognition domain also comes from an unusual class of mutants¹⁹, isolated using a genetic screen to identify amino-acid substitutions that enable *FokI* to cleave DNA in cells expressing the cognate methylase activity. When applied to the restriction endonuclease *EcoRI*, a similar screen produced mutants that cleaved DNA at non-cognate sites and the mutations occurred in residues involved in DNA recognition²⁰. In the *FokI* screen, eight mutations (G188K, P196S, T343I, S388N, S395F, E407K, E410K, D421N) were reported and two (P196S, D421N) were studied in detail¹⁹. Unlike *EcoRI*, both of these mutants cleaved only at the cognate *FokI* site, including hemimethylated sites in which one or the other DNA strand was methylated. Remarkably, we find that seven of the eight mutations occur at the interdomain interface and

would be expected to relax the domain–domain association (Fig. 5). The mutants T343I, S388N and S395F would disrupt the hydrogen-bonding network between residues Ser 395, Lys 387 and Glu 339 that connects the cleavage domain to D3. Mutants G188K, E407K and E410K insert a positive charge in the vicinity of residue Lys 427, destabilizing the electrostatic environment at the interface. Replacement of Pro 196 with a serine residue will weaken a key hydrophobic interaction between D2 and helix $\alpha 3$ of the cleavage domain, and may also disrupt a nearby hydrogen bond between Ser 194 and Lys 427. To explain their data, Waugh and Sauer postulated that there may be a rate-limiting step in the *FokI* cleavage reaction that involves the dissociation of the cleavage domain from the recognition domain which is accelerated by these mutants to allow cleavages of even hemimethylated DNA sites¹⁹. The compact nature of our complex and the mapping of these mutations to the interdomain interface is consistent with this new mechanism of enzyme activation.

The *FokI* complex has implications for the evolution of HTH proteins. The *FokI* structure reveals unprecedented modifications to the basic CAP core, including the insertion of large loops and secondary structural elements in subdomains D1 and D2. These extra elements act as additional prongs both to grip the DNA and to mediate protein–protein interactions between the subdomains. These embellishments are so extensive that the structural alignment program DALI (ref. 21) was needed to elucidate the CAP core. The D3 recognition helix lies outside the major groove, suggesting that the HTH motif can perform other roles besides DNA recognition. An analogous mechanism can be found in multi-zinc-finger proteins such as TFIIIA^{22,23} and GLI²⁴, in which some zinc-fingers penetrate the DNA major groove to recognize specific base pairs, and others lie outside the major groove to fulfil other functions. The evolutionary success of the HTH motif in both prokaryotes and eukaryotes appears to be due partly to its versatility in accommodating modifications both in structure and in function.

Finally, the *FokI* complex provides a framework for the design of

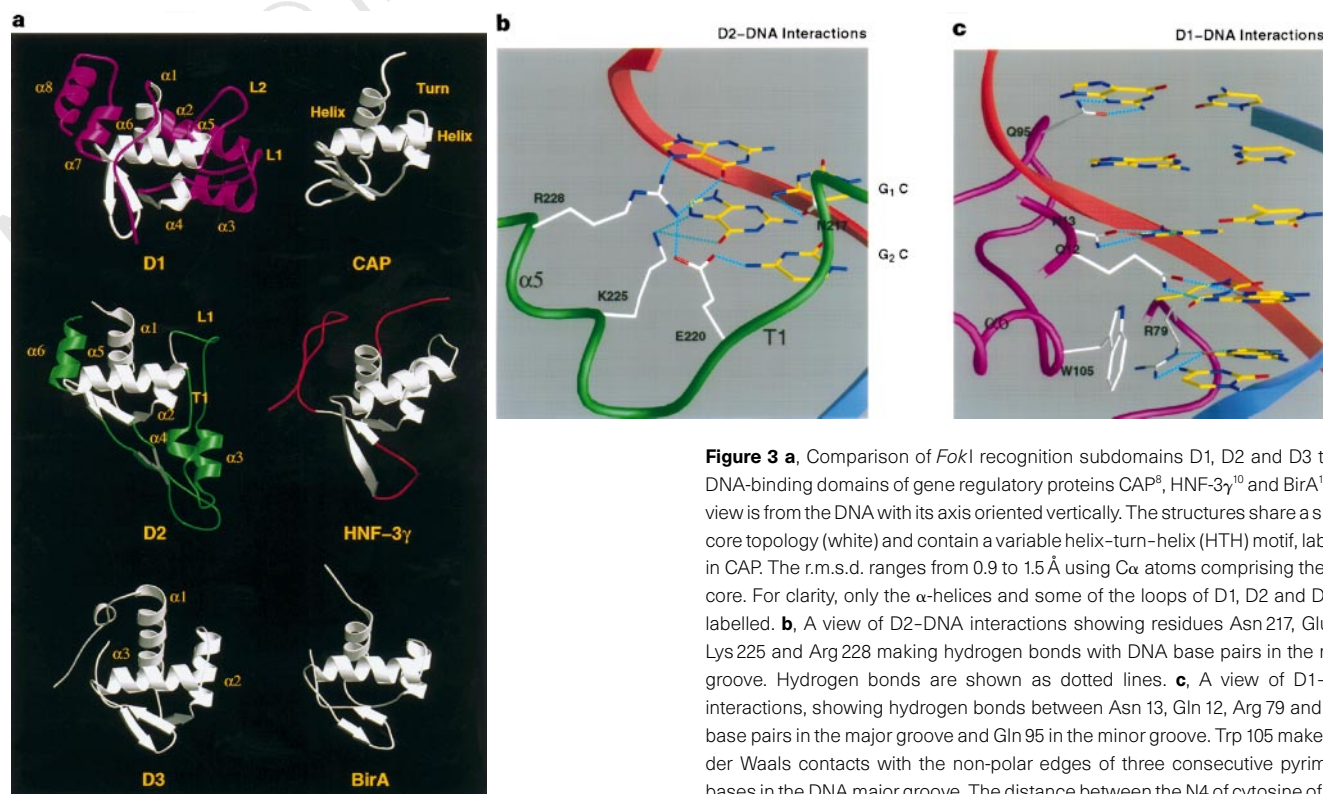


Figure 3 **a**, Comparison of *FokI* recognition subdomains D1, D2 and D3 to the DNA-binding domains of gene regulatory proteins CAP⁸, HNF-3 γ ¹⁰ and BirA¹¹. The view is from the DNA with its axis oriented vertically. The structures share a similar core topology (white) and contain a variable helix–turn–helix (HTH) motif, labelled in CAP. The r.m.s.d. ranges from 0.9 to 1.5 Å using C α atoms comprising the CAP core. For clarity, only the α -helices and some of the loops of D1, D2 and D3 are labelled. **b**, A view of D2–DNA interactions showing residues Asn 217, Glu 220, Lys 225 and Arg 228 making hydrogen bonds with DNA base pairs in the major groove. Hydrogen bonds are shown as dotted lines. **c**, A view of D1–DNA interactions, showing hydrogen bonds between Asn 13, Gln 12, Arg 79 and DNA base pairs in the major groove and Gln 95 in the minor groove. Trp 105 makes van der Waals contacts with the non-polar edges of three consecutive pyrimidine bases in the DNA major groove. The distance between the N4 of cytosine of base pair G₅C and the tryptophan indole ring is 3.9 Å, ruling out a strong aromatic hydrogen bond.

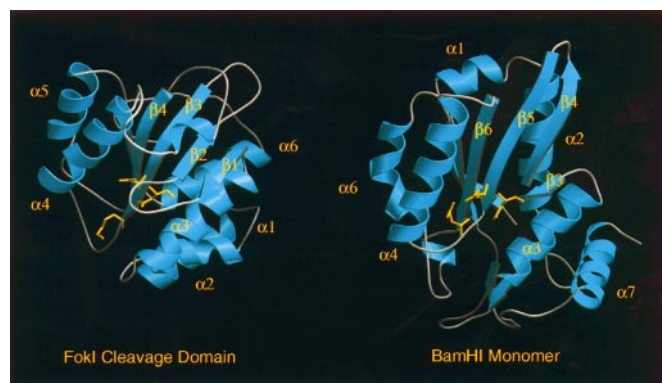


Figure 4 Comparison of *FokI* cleavage domain to a monomer of endonuclease *BamHI*. The two structures share a similar β -sheet core surrounded by α -helices on both sides, and superimpose with an r.m.s.d. of 2.0 Å using 61 C α atoms. The active-site residues of both nucleases occur at one end of the β -sheet and are shown in yellow.

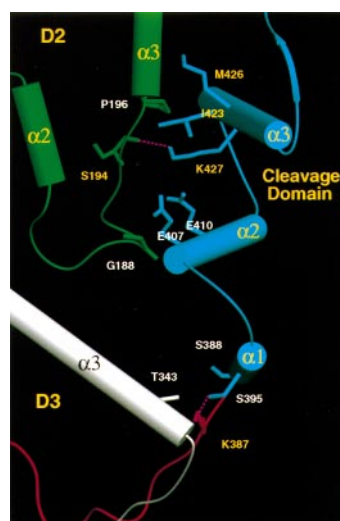


Figure 5 The new class of *FokI* mutants (labelled in white) identified by Waugh and Sauer¹⁹ map to the interdomain interface between the recognition domain and the cleavage domain. The mutations appear to disrupt the interdomain interface, allowing the cleavage domain more readily to dissociate from the recognition domain and swing over to the DNA major groove for cleavage.

new restriction enzymes. Chimaeric enzymes can be created by attaching the cleavage domain of *FokI* to the DNA-binding domains of transcription factors. The structure allows a definition of the linker segment precisely (residues 373 to 387) and allows a modelling of the *FokI* cleavage domain alongside the DNA-binding domains of various transcription factors. The zinc-finger transcription factors Zif268 and Sp1 have received particular attention owing to the possibility of selecting and swapping fingers to create new specificities^{5,6}. The mechanism of nuclease activation implied by our structure may be an additional feature that could be incorporated into the design of artificial enzymes as a way of regulating their activity. □

Methods

FokI was expressed, purified and co-crystallized with a 20-bp DNA oligonucleotide in 2.2 M ammonium sulphate²⁵. The co-crystals diffract to 2.8 Å when cooled to -160°C and exposed to synchrotron radiation. They belong to space group $P2_1$ with unit cell dimensions of $a = 65.6$ Å, $b = 119.3$ Å, $c = 71.5$ Å, and $\beta = 101.4^{\circ}$, and contain a single protein–DNA complex per asymmetric unit. Data were measured at the Cornell High Energy Synchrotron Source (CHESS), using both the charge-coupled device (CCD) detector (beamline A1) and imaging plates (beamline F1) at $\lambda = 0.908$ Å (Table 1). Most of the heavy-atom derivatives were prepared by conventional soaks in different heavy-atom solutions. The Iodo derivative was prepared by substituting iodouracils for four thymine residues on the DNA fragment (Fig. 1). The

heavy-atom positions were refined with PHASIT²⁶ and the resulting MIRAS (multiple isomorphous replacement including anomalous scattering) phases were used to calculate a 3.5 Å map that was continuous and showed clear electron density for the DNA. The position of the DNA matched the iodouracil positions determined from the Iodo derivative. The MIRAS map was further improved by cycles of solvent flattening with SOLOMON²⁷. Using these and subsequent partial model phase combined maps, interspersed with positional and simulated annealing refinements using X-PLOR²⁸, we built 562 residues of the protein and the complete DNA molecule with program O²⁹. Residues 1–3, 252–254 and 282–286 could not be built owing to uninterpretable density. The positional and temperature factor refinements were extended to 2.8 Å resolution²⁸, and the model was verified through extensive simulated annealing omit maps. Near the end of refinement, 172 water molecules were added to the model (B -values <65 Å²) from the inspection of $F_o - F_c$ maps.

Received 23 December 1996; accepted 23 April 1997.

1. Sugisaki, H. & Kanazawa, S. New restriction endonucleases from *Flavobacterium okeanokoites* (*FokI*) and *Micrococcus luteus* (*MluI*). *Gene* **16**, 73–78 (1981).
2. Li, L., Wu, L. P. & Chandrasegaran, S. Functional domains of *FokI* restriction endonuclease. *Proc. Natl Acad. Sci. USA* **89**, 4275–4279 (1992).
3. Skowron, P., Kaczorowski, T., Tucholski, J. & Podhajska, A. J. Atypical DNA-binding properties of class-IIIS restriction endonucleases: evidence for recognition of the cognate sequence by a *FokI* monomer. *Gene* **125**, 1–10 (1993).
4. Kim, Y.-G. & Chandrasegaran, S. Chimeric restriction endonuclease. *Proc. Natl Acad. Sci. USA* **91**, 883–887 (1994).
5. Kim, Y.-G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: Zinc finger fusions to *FokI* cleavage domain. *Proc. Natl Acad. Sci. USA* **93**, 1156–1160 (1996).
6. Huang, B., Schaeffer, C. J., Li, Q. & Tsai, M.-D. Splase: a new class IIIS zinc-finger restriction endonuclease with specificity for SP1 binding sites. *J. Prot. Chem.* **15**, 481–489 (1996).
7. Podhajska, A. J. & Szybalski, W. Conversion of the *FokI* endonuclease to a universal restriction enzyme: cleavage of phase M13mp7 DNA at predetermined sites. *Gene* **40**, 175–182 (1985).
8. Schultz, S. C., Shields, G. C. & Steitz, T. A. Crystal structure of a CAP–DNA complex: the DNA is bent by 90 degrees. *Science* **253**, 1001–1007 (1991).
9. Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L. & Sweet, R. M. Crystal structure of globular domain histone H5 and its implications for nucleosome binding. *Nature* **362**, 219–223 (1993).
10. Clark, K. L., Halay, E. D., Lai, E. & Burley, S. K. Co-crystal structure of the HNF-3 γ DNA recognition motif resembles histone H5. *Nature* **364**, 412–420 (1993).
11. Wilson, K. S., Shewchuk, L. M., Brennan, R. G., Otsuka, A. J. & Matthews, B. W. *Escherichia coli* biotin holoenzyme synthetase/biorepressor crystal structure delineates biotin- and DNA-binding domains. *Proc. Natl Acad. Sci. USA* **89**, 9257–9261 (1992).
12. Harrison, S. C. & Aggarwal, A. K. DNA recognition by proteins with the helix–turn–helix motif. *Annu. Rev. Biochem.* **59**, 933–969 (1990).
13. Pavletich, N. P. & Pabo, C. O. Zinc finger–DNA recognition: crystal structure of a zif268–DNA complex at 2.1 Å resolution. *Science* **252**, 809–817 (1991).
14. Klemm, J. D., Rould, M. A., Aurora, R., Herr, W. & Pabo, C. O. Crystal structure of Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell* **77**, 21–32 (1994).
15. Newman, M., Strzelecka, T., Dörner, L. F., Schildkraut, I. & Aggarwal, A. K. Structure of restriction endonuclease *BamHI* and its relationship to *EcoRI*. *Nature* **268**, 660–664 (1994).
16. Waugh, D. S. & Sauer, R. T. Single amino acid substitutions uncouple the DNA binding and strand scission activities of *FokI* endonuclease. *Proc. Natl Acad. Sci. USA* **90**, 9596–9600 (1993).
17. Yonezawa, A. & Sugiura, Y. DNA binding mode of class-IIIS restriction endonuclease *FokI* revealed by DNA footprinting analysis. *Biochem. Biophys. Acta* **1219**, 369–379 (1994).
18. Li, L., Wu, L. P., Clarke, R. & Chandrasegaran, S. C-terminal deletion mutants of the *FokI* restriction endonuclease. *Gene* **133**, 79–84 (1993).
19. Waugh, D. S. & Sauer, R. T. A novel class of *FokI* restriction endonuclease mutants that cleave hemimethylated substrates. *J. Biol. Chem.* **269**, 12298–12303 (1994).
20. Heitman, J. & Model, P. Mutants of the *EcoRI* endonuclease with promiscuous substrate specificity implicate residues involved in substrate recognition. *EMBO J.* **9**, 3369–3378 (1990).
21. Holm, L. & Sander, C. J. Protein structural comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
22. Churchill, M. E., Tullius, T. D. & Klug, A. Mode of interaction of the zinc finger protein TFIIIA with a 5S RNA gene of *Xenopus*. *Proc. Natl Acad. Sci. USA* **87**, 5528–5532 (1990).
23. Clemens, K. R., Liao, X., Wolf, V., Wright, P. E. & Gottesfeld, J. M. Definition of the binding sites of individual zinc fingers in the transcription factor IIIA–5S RNA gene complex. *Proc. Natl Acad. Sci. USA* **89**, 10822–10826 (1992).
24. Pavletich, N. P. & Pabo, C. O. Crystal structure of a five-finger GLI–DNA complex: new perspectives on Zn fingers. *Science* **261**, 1701–1707 (1993).
25. Hirsch, J. A., Wah, D. A., Dörner, L. F., Schildkraut, I. & Aggarwal, A. K. Crystallization and preliminary X-ray analysis of restriction endonuclease *FokI* bound to DNA. *FEBS Lett.* **403**, 136–138 (1997).
26. Furey, W. *Phases—A Program Package for the Processing and Analysis of Diffraction Data for Macromolecules* (VA Medical Center, Pittsburgh, PA, 1993).
27. Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F₁ ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
28. Brunger, A. T. *X-PLOR, Version 3.1, A System for X-ray Crystallography and NMR* (Yale University, New Haven, CT, 1992).
29. Jones, T. A., Zou, J.-Y. & Cowan, S. W. Improved methods for building models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).

Acknowledgements. We thank the staff at CHESS and C. Escalante, E. Jacobson, M. Newman and H. Viadiu for help with data collection; R. Knott for DNA synthesis; and T. Bestor, W. Hendrickson, R. Mann and R. Roberts for comments on the manuscript. This work was supported by a NIH grant (A.K.A.) and NIH training grants (J.A.H. and D.A.W.).

Correspondence and requests for materials should be addressed to A.K.A. (e-mail: aggarwal@inka.mssm.edu). Coordinates have been deposited in the Brookhaven Protein Data Bank under accession number 1FOK.

Taking inventory of biological reagents

Whether you are combing the shelves or digging in the back of the freezer, that missing reagent might just be here, such as cell-cycle antibodies, peptide controls, CO and haem oxygenases and those elusive FISH probes.

Bioreagents

From Molecular Biosciences

A variety of bioligands, crosslinking reagents and fluorescent probes used to investigate mechanisms of cellular function

The company produces over 200 high-purity research tools, such as water-soluble biotinylation reagents, reactive cholesterol, vitamin derivatives, lipophilic probes, fluorophores, radioiodination reagents and dendritic cores for preparing oligomers. These materials are used to modify polypeptides, oligonucleotides, monoclonal antibodies, as well as other biomolecules and surfaces. The company is developing new products and technology for diagnostics, drug delivery and pharmaceutical research products.

Reader Enquiry No. 100

Protease inhibition reference

From Boehringer Mannheim Biochemicals

A new technical brochure to guide in the selection of protease inhibition compounds

This guide provides detailed information about when and how to use protease inhibitors. It features the company's complete protease inhibitor cocktail tablets (EDTA-free and mini-tablets), which are stated to offer greater protection in aqueous buffers. It also describes the new Pefabloc SC PLUS and PMSF PLUS sets. These include reagents that inhibit possible covalent protein/protease inhibitor interactions, without influencing the effectiveness of protease inhibition.

Reader Enquiry No. 101

Actin polymerization biochem kit

From Cytoskeleton

Contains all the components necessary for researchers to determine the effects of actin binding proteins on actin polymerization.

The kit allows the user to differentiate between nucleating, capping, side binding and severing proteins. Twenty additional biochem kits are available from Cytoskeleton to study aspects of tubulin, actin and intermediate filament function.

Reader Enquiry No. 102

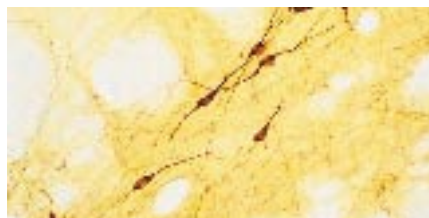
Immunology and histology sources

Peptide controls

From Incstar

A number of peptide controls for use with the company's histochemical antisera

Peptide controls are now available for the recently released histochemical antisera against 5-HT transporter, 5-HT_{2A} receptor,



Labelling of rat striatum with bNOS antiserum.

brain nitric oxide synthase, C-FOS, vesicular acetylcholine transporter, μ -opioid receptor and δ -opioid receptor. These synthetic peptides are intended to be used as adsorption controls for their respective antibodies and are crucial for the validation of specific staining results. Pre-adsorption with the specific peptide immunogen should provide complete blockage of stain.

Reader Enquiry No. 103

Antibody Source

From Biodesign

Over 575 new product additions to the company's existing range of products

The 1997-98 catalogue provides details on some 4,000 immunological reagents specific to apolipoproteins, apoptosis, autoimmune, cancer antigens, cardiac markers, cell receptors, collagens and conjugates. Additional product categories include cytokines, hormones, immunoassay development reagents, infectious diseases, immunoglobulins, signal transduction, TMB substrate and tumour markers.

Reader Enquiry No. 104

Cell-cycle antibodies

From TCS Biologicals

A group of cyclin-cdk binding proteins has recently been characterized that inhibit the activity of serine/threonine kinases (cdks)

According to the manufacturer, p21^{kip1} may be a universal inhibitor of cyclin-dependent kinases because it can inhibit the activity of each member of the cyclin-cdk family. It can block the phosphorylation of the Rb gene product and is an important downstream effector of p53-mediated growth control of mammalian cells.

Another of these cyclin-cdk inhibitors, p16^{ink4}, has been shown to bind strongly to the cyclin D/cdk4 complex; p16^{ink4} is either deleted or rearranged in many human cell lines derived from tumours. p27^{kip1} binds strongly to the cyclin E/cdk2 complex in mink lung cells arrested by either TGF- β or contact inhibitor. The cyclin-cdk inhibitor proteins, p21^{kip1}, p16^{ink4} and

p27^{kip1} may mediate the action of tumour suppressor gene products.

Reader Enquiry No. 105

Human IgG subclasses

From CLB

A 52-page booklet on human IgG subclasses as diagnostic markers

The structural and functional differences between the four subclasses are outlined, as well as their differences in reactivity with Fc receptors and their role in humoral immune responses. The clinical relevance of abnormal levels of these proteins is dealt with in relation to a broad spectrum of immune-mediated disease states, with a focus on the question: Under which circumstances should IgG subclass levels should be determined and what conclusions may be drawn from such measurements? With regard to technical aspects, the six most widely used techniques for the assay of the IgG subclasses are outlined, together with their advantages and disadvantages. The booklet is a useful source of information, is suitable as a quick reference in this field and is available free of charge.

Reader Enquiry No. 106

CD95, Fas clone APO-1 mAbs

From Dako

Monoclonal antibody against CD95, Fas clone APO-1 antibody for use in immunohistochemistry or flow cytometry applications

CD95 is a transmembrane protein that is known to trigger apoptosis in a variety of cell types upon anti-APO-1 binding or binding of the Fas ligand. As such, CD95 is a protein that is closely associated with immunological cell death. For immunohistochemistry, the company offers an optimized procedure for routinely processed tissue sections. The company's LSAB+ high-performance detection system is complemented by this procedure. Reproducible flow cytometry analyses are also obtained with the procedure. Other apoptosis-related antibodies are available from the company and include bc1-2, p53, Lewis^y, NGFR and CD95 (APO-1/Fas).

Reader Enquiry No. 107



Dako's mAb to CD95, Fas Clone APO-1.

Atlas of HEp-2 Patterns

From The Binding Site

Recently published is the 'Atlas of autoantibody patterns on tissues'

The Atlas is intended as both an educational guide and a reference text for laboratory staff who have the task of interpreting immuno-fluorescent autoantibody patterns and informing clinicians of their clinical relevance. Many common autoantibody patterns are described, together with clinical associations, autoantigens and references. Over 80 high-quality colour photographs have been chosen to illustrate typical patterns on rodent and primate tissue and guidance is given on potential identification problems.

Reader Enquiry No. 108

Neuroscience reagents

CO and haem oxygenases

From Affiniti Research Products

A new range of reagents to help understand carbon monoxide and haem oxygenases in physiology and pathophysiology, respectively

The catabolism of haem to biliverdin with the release of iron and carbon monoxide (CO) is mediated by the enzyme haem oxygenase (HO). Two distinct forms of haem oxygenase (HO-1-inducible and HO-2-constitutive) have been characterized from mammals and may prove useful to researchers investigating the role of CO as a second messenger molecule and putative neurotransmitter. Using the synthetic peptide approach to antiserum production, polyclonal antisera wholly specific to each HO isoform have been generated and characterized for use in western blotting and immunohistochemical applications. In addition to antisera, Affiniti's HO range also includes HO-enriched microsomal preparations of several rat tissues, which serve as pos-

itive control lysates for western blotting. Complete abolition of antiserum reactivity may be obtained when preadsorption with the synthetic peptides used to generate the antisera is carried out. Both peptides are offered as negative controls and as solid-phase ligands. Finally, a large range of HO substrates, inducers and inhibitors, including several porphyrin compounds is available.

Reader Enquiry No. 109

SSRI

From RBI

Fluoxetine hydrochloride and its metabolite, norfluoxetine hydrochloride are now offered strictly for non-human research use

Commonly known as Prozac, fluoxetine was the first selective serotonin reuptake inhibitor (SSRI) approved for use in humans as an antidepressant, and has become the most widely prescribed drug in its class. These compounds will be of interest to researchers studying the role of serotonin in brain function and, in particular, in the areas of depression, aggression, eating disorders and sexual dysfunction. Fluoxetine hydrochloride, is available in 50-mg and 10-mg units and norfluoxetine hydrochloride is available in 2-mg units.

Reader Enquiry No. 110

DNA FISH probes

From Alpha Laboratories

A high-quality range of DNA probes for research into haematologic cancers together with DNA probe kits for IVD use

These Vysis probes are directly labelled, offering rapid, easy-to-use protocols that provide bright, clear results with little background staining. Clarity is one of the main factors that should make these probes useful for identifying new genetic anomalies associated

with haematologic cancers and for monitoring patients. The extensive range of probes includes the highly successful bcr/abl translocation probe and the newly released 15;17 translocation probe that is so important in the characterization of APL. Other products include CEP 8 for enumeration of chromosome 8 and CEP X/Y for evaluation of engraftment following sex mismatched bone marrow transplantation.

Reader Enquiry No. 111

PathDetect

From Stratagene

Rapid, convenient in vivo signal transduction pathway reporting systems

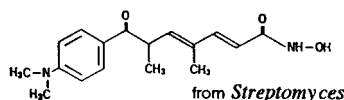
These reporting systems can be used to study the *in vivo* effects of new genes, growth factors and drug candidates on the activation of c-Jun N-terminal kinase (JNK), mitogen-activated protein kinase (MAPK), cyclic AMP-dependent protein (PKA) and other signalling molecules leading to the activation of these kinases. When a fusion activator plasmid, reporter plasmid and uncharacterized gene are co-transfected into mammalian cells, either direct or indirect phosphorylation of the fusion activator protein by the uncharacterized gene product will activate transcription of the luciferase gene from the reporter plasmid. Also the effects of extracellular stimuli can be studied. According to the manufacturer, activation as high as 200-fold with very low background has been observed with the Path-Detect c-Jun, Elk1 and CREB trans-reporting systems.

Reader Enquiry No. 112

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

Trichostatin A



Catalog No. 204-11991

This histone deacetylase inhibitor is highly effective at low concentrations, with high specificity. For more information, please call
Wako BioProducts: (800) 992-9256; (804) 271-7677.

READER ENQUIRY NO. 98

CUSTOM HYBRIDOMA DEVELOPMENT

- Ascites Production
 - Purification • Conjugation
 - Polyclonal Antibodies
- Also Available

1-800-481-9737

e-mail: antibodies@htbio.com

http://www.htbio.com

fax: (760) 788-9694

P.O. Box 1319

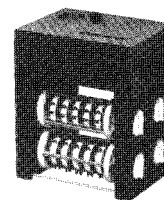
Ramona, CA

92065

HTI
 HTI Bio-Products, Inc.

READER ENQUIRY NO. 97

BIANORM® EQUILIBRIUM DIALYSIS SYSTEM



The gentlest method to quantify the binding parameters and thermodynamic data between ligands and biopolymers. Results come closest to the actual values. Sample volume from 0.2 ml to 5.0 ml. Up to 20 experiments in parallel.

BIANORM
 Geräte
 P.O. Box 65012
 D-81215 Munich

READER ENQUIRY NO. 61